

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

A Thesis Submitted for the Degree of PhD at the University of Warwick

<http://go.warwick.ac.uk/wrap/34663>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

The impact of uninsurable risk on asset prices and optimal dividend policy.

Mark Charles Freeman

August 1997

A thesis presented for the degree of Doctor of
Philosophy for the University of Warwick

Warwick Business School
University of Warwick.

Summary

This thesis examines whether two well documented financial market anomalies — the “Mehra & Prescott puzzles” and the dividend controversy — can be resolved by allowing for the effects of uninsurable risks. The dissertation contains an extensive review of the theory of consumption based asset pricing and the Mehra & Prescott puzzles. This provides more comprehensive coverage of this material than any previous review of the area: see chapters 2 & 3. The role that uninsurable risk might play in resolving market anomalies is clearly demonstrated. Three chapters of substantive original contribution follow that examine: (i) the predicted equity premium when marketable and nonmarketable risks are independent (ii) the potential relevance of aggregate dividends to equilibrium asset prices in economies with idiosyncratic endowment shocks and (iii) the response of the stock market and riskfree rate to unemployment shocks. The main findings are: (i) Chapter 4: an integrated approach to local proper risk aversion is presented and a new form of risk aversion emerges naturally (ii) Chapter 4: it will not, in general, be possible to make accurate quantitative predictions concerning the impact of a small probability, high impact, negative shock to endowment (“unemployment”) on asset prices on the basis of current knowledge concerning investor preferences (iii) Chapter 5: aggregate dividends are shown to play an important role in helping individuals to consumption smooth in incomplete markets if the level of aggregate investment is uncertain. The observed behaviour of dividend smoothing and concentrating rights issues into times of economic prosperity is consistent with the model that is presented (iv) Chapter 6: the rise (fall) in the riskfree rate prior to “bad” (“good”) unemployment news does not appear to be consistent with precautionary savings behaviour. It is concluded that, while incomplete market models have great theoretical strength and some empirical support, current applications of this theory leave many issues unresolved.

Contents

| | | |
|-----------|---|-----------|
| I | Introduction | 17 |
| 1 | Incomplete markets, financial puzzles | 18 |
| 1.1 | Issues to be addressed | 18 |
| 1.2 | Sources of nontradable risks. | 21 |
| 1.3 | Market anomalies and incompleteness | 25 |
| 1.4 | The development of this thesis | 29 |
| II | Consumption based asset pricing | 36 |
| 2 | Introduction | 38 |
| 3 | Fundamental theorem of asset pricing | 42 |
| 3.1 | Introduction | 42 |
| 3.2 | The fundamental theorem in economies with finite states . . . | 44 |
| 3.3 | A Hilbert space proof of the fundamental theorem | 49 |
| 4 | Investor preferences | 53 |
| 5 | The Euler equation and discrete time theory | 59 |
| 5.1 | Complete markets and aggregate consumption | 60 |
| 5.2 | The representative agent | 63 |
| 5.3 | The Euler equation for the representative agent | 64 |
| 5.4 | The Euler equation in incomplete markets | 66 |

| | | |
|----------|--|-----------|
| 5.5 | Towards the CCAPM | 68 |
| 5.6 | The CCAPM in discrete time | 70 |
| 5.6.1 | The CCAPM under bivariate normality | 70 |
| 5.6.2 | The CCAPM under quadratic utility | 71 |
| 5.6.3 | The CCAPM in discrete time under isoelastic utility, lognormal consumption & normal asset returns | 72 |
| 5.6.4 | The CCAPM: an approximation under isoelastic utility and smooth consumption | 72 |
| 6 | The CCAPM in continuous time | 73 |
| 6.1 | The Intertemporal Capital Asset Pricing Model | 77 |
| 6.2 | The Consumption Capital Asset Pricing Model | 80 |
| 6.2.1 | The Breeden proof | 80 |
| 6.2.2 | Exploiting local linearity | 82 |
| 6.3 | The CCAPM and ICAPM with nonmarketable assets | 83 |
| 6.4 | CAPM versus CCAPM | 88 |
| 7 | Developing the Mehra & Prescott puzzles | 91 |
| 7.1 | A one-period economy | 92 |
| 7.1.1 | Non-parametric | 92 |
| 7.1.2 | A parametric approach | 92 |
| 7.2 | A Merton (1971) style analysis | 93 |
| 7.3 | Markov growth model | 95 |
| 7.4 | Comparing the models | 96 |

| | |
|--|------------|
| 8 Conclusion | 97 |
| 9 Appendix | 99 |
| 9.1 Probability space | 99 |
| 9.2 Vector spaces — Definitions | 102 |
| 9.3 Riesz Representation Theorem | 104 |
| 9.4 Applications of the fundamental theorem | 105 |
| 9.4.1 The Capital Asset Pricing Model | 106 |
| 9.4.2 Arbitrage Pricing Theory | 107 |
| 9.4.3 Binomial Options Pricing | 108 |
| 9.5 Dynamic programming | 111 |
| III The Mehra & Prescott puzzles | 114 |
| 10 Introduction | 116 |
| 11 Developing the Mehra & Prescott puzzles | 120 |
| 11.1 Background | 120 |
| 11.2 Formal development | 124 |
| 12 Possible explanations of the equity premium puzzle | 128 |
| 12.1 Are these really puzzles? | 131 |
| 12.2 Form of the investor's utility function | 138 |
| 13 The Mehra & Prescott puzzles in incomplete markets | 142 |
| 13.1 Introduction | 142 |

| | | |
|-----------|--|------------|
| 13.2 | Precautionary savings and liquidity constraints | 145 |
| 13.2.1 | The theoretical approaches | 148 |
| 13.2.2 | Empirical evidence | 153 |
| 13.3 | Risky and riskless assets | 156 |
| 13.3.1 | A one period model | 158 |
| 13.3.2 | Multiperiod models | 160 |
| 13.4 | Savings behaviour of the unemployed | 166 |
| 14 | Closely related puzzles | 170 |
| 14.1 | Hansen-Jagannathan bounds | 171 |
| 14.2 | (Co)Variances of asset returns | 178 |
| 15 | Empirical tests of consumption based asset pricing models | 179 |
| 15.1 | Parametric tests | 181 |
| 15.2 | Non-parametric tests | 184 |
| 15.3 | Realised returns vs. consumption betas | 187 |
| 15.4 | Adjusting spot consumption data | 188 |
| 16 | Conclusion | 189 |
| IV | Proper risk aversion and the Mehra & Prescott puzzles | 194 |
| 17 | Introduction | 196 |
| 18 | Proper risk aversion | 198 |

| | | |
|-----------|---|------------|
| 18.1 | On the utility function | 199 |
| 18.1.1 | Risk Aversion, Prudence and Temperance | 199 |
| 18.1.2 | Risk aversion with stochastic consumption | 202 |
| 18.2 | Standard and Proper Risk Aversions | 204 |
| 18.2.1 | Local proper risk aversion | 206 |
| 18.2.2 | Global proper risk aversion | 213 |
| 19 | Weil's example | 215 |
| 19.1 | The issue to be addressed | 218 |
| 20 | The magnitude of the equity premium | 223 |
| 20.1 | Creating alternate utility functions | 223 |
| 20.2 | Results and discussion | 228 |
| 21 | Conclusion | 232 |
| V | Uninsurable risk and optimal dividend policy | 236 |
| 22 | Introduction | 238 |
| 22.1 | The economy | 241 |
| 22.2 | The key assumptions | 245 |
| 22.3 | The theory of dividend policy | 249 |
| 22.4 | The empirical evidence | 251 |
| 23 | A model with trading | 256 |
| 23.1 | Describing the economy | 256 |

| | | |
|-----------|---|------------|
| 23.2 | No trade | 261 |
| 23.2.1 | Is a market required? | 262 |
| 23.2.2 | Optimal managerial policy | 263 |
| 23.3 | With a financial market | 266 |
| 23.3.1 | Certain returns | 266 |
| 23.3.2 | State independent risky returns | 270 |
| 23.3.3 | State dependent risky returns | 272 |
| 24 | A multi-period economy | 274 |
| 24.1 | Certain and state dependent risky returns | 275 |
| 24.2 | A dividend smoothing model | 281 |
| 25 | Conclusion | 284 |
| 26 | Appendix | 286 |
| VI | Unemployment shocks and asset returns | 293 |
| 27 | Introduction | 295 |
| 28 | The economic environment | 300 |
| 28.1 | A one period model | 300 |
| 28.2 | Application of model to real data | 303 |
| 28.3 | Unemployment shocks and asset returns | 306 |
| 29 | Empirical testing | 308 |

| | |
|--|----------------|
| 29.1 Data | 308 |
| 29.2 Regression analysis | 312 |
| 29.3 Further empirical investigation | 316 |
| 30 Conclusion | 323 |
| VII Conclusion | 326 |
| 31 Conclusion | 327 |
| 31.1 Summary of findings | 327 |
| 31.2 Further research | 333 |
| 31.3 Final comments | 340 |
| VIII Bibliography | 343 |

List of Figures

- 1 Quarterly real aggregate consumption on non-durables in the UK and the real value of the All-share index with dividends reinvested 1967Q1 – 1992Q2. Both series are readjusted so that the 1985Q1 value is 1. 124
- 2 The riskfree rate / equity premium bound (below the line). The top graph is given by a simple one period model where consumption is assumed to be smooth. The bottom graph is given by a one period model in the style of Hansen and Singleton where consumption growth and the return to the market are assumed to be lognormal. 129
- 3 The riskfree rate / equity premium bound. The top graph is given by a Markov-growth model. This figure can be directly compared with figure 1 in Mehra & Prescott (1985). The bottom graph is given by a continuous-time model in the spirit of Merton (1971). 130
- 4 Two state economy, income & dividend risk related, size of risk needed to explain the Mehra & Prescott puzzles. 135
- 5 Two state economy, income risk concentrated in a subset of the population, size of risk needed to explain the Mehra & Prescott puzzles. 160

| | | |
|---|---|-----|
| 6 | The Hansen-Jagannathan bound tested on annual consumption data from the US 1890-1980: source Grossman, Melino & Shiller (1987). The “CRRA” points give the mean and standard deviation of the IMRS using this consumption data with a CRRA utility function for $\beta = 0.97, \gamma \in [1, 25]$. As γ rises, so the standard deviation rises. The two solid lines give the Hansen-Jagannathan bound for this mean / standard deviation pair. The Koch(90) line uses the average real returns and (co)variances to the market and a riskfree asset for the US over the interval 1888–1978 as given by Kocherlakota (1990b). The GMS(87) line uses the real returns and (co)variances to the market, a short term riskfree asset and a long term bond index over the interval 1890–1980 as given by Grossman, Melino & Shiller (1987). | 174 |
| 7 | Sufficient and necessary & sufficient conditions for the various forms of local and global proper risk aversion. | 214 |
| 8 | These graphs replicate and adjust two of the three graphs on page 786 of Weil (1992a). Unemployed income is one third the employed income. The top (bottom) graph is the ratios of the predicted equity premium (riskfree rate) in the presence of unemployment risk to the predicted equity premium (riskfree rate) in the absence of unemployment risk. | 231 |

9 These graphs replicate and adjust two of the three graphs on page 787 of Weil (1992a). Unemployed income is zero. The top (bottom) graph is the ratios of the predicted equity premium (riskfree rate) in the presence of unemployment risk to the predicted equity premium (riskfree rate) in the absence of unemployment risk. 233

10 Top graph: fixed investment by manufacturing companies including leasing in 1990 in the UK (left hand axis) and the nominal riskfree rate (right hand axis) 1995Q1 – 1993Q4. Source: Office of National Statistics and the LSPD. Bottom graph: the number of dividend omissions against unemployment for the US 1963–86. Data on dividend omissions is described in Sant and Cowan (1994) and unemployment data is discussed in the next chapter. Dividend omissions should be read against the right hand axis. The max-min bars, which should be read against the left hand axis, give the range of monthly unemployment over the year. The Unempl. line gives average monthly unemployment in that year. 254

- 11 A graph showing how the significance of state dependent investment uncertainty varies with unemployment risk, unemployment income, risk aversion and fixed return on capital for power utility. M measures the additional percentage of shares investors will choose to take up at $t = 0$ if investment is uncertain in the high state rather than the low state. γ is the coefficient of relative risk aversion, r is the return on capital (non-stochastic). s is the unemployment rate in the low state and y_u is endowment in the unemployed state. For all iterations, $i = 0.1, y_h = 1$. For the top graph $\gamma = 3, r = 0.1$. For the bottom graph, $s = 0.05, y_u = 0.4$ 271
- 12 A graph showing how the significance of state dependent investment uncertainty varies with “profrrsk” $\equiv \vartheta$, which is half the difference between returns in the low state and returns in the high state and “retained” $\equiv \nu$, which is the amount of investment funded from retained profits. M measures the additional percentage of shares that investors will choose to take up at $t = 0$ if investment uncertainty is in the high state rather than the low state. For all iterations, $i = 0.15, \gamma = 3, y_u = 0.4, s = 0.05, y_h = 1, E[r] = 0.1$ 279

- 13 Cumulative abnormal returns to the equity premium around a shock in unemployment as modelled by S_t . Here “abnormal return” at month t is the difference between the equity premium in that month and its average over the whole sample: $p_{ij}, i = -12, j \in [-12 + 6]$. Unemployment shocks are divided into equally sized quartiles. Month 0 is the month referred to in the unemployment statistic. The sample period is June 1971 to March 1993 and is based on UK data. 318
- 14 Cumulative changes in the nominal riskfree rate around a shock in unemployment as modelled by S_t : $f_{ij}, i = -12, j \in [-12, +6]$. 0 on the y-axis refers to the riskfree rate in month -13. Unemployment shocks are divided into equally sized quartiles. Month 0 is the month referred to in the unemployment statistic. The sample period is June 1971 to March 1993 and is based on UK data. 319

List of Tables

| | | |
|---|--|-----|
| 1 | Representing the linear relationship between prices and pay-outs as a positive pricing kernel and equivalent martingale measure (EMM). | 50 |
| 2 | Evidence on the first and second moments of the equity premium and real riskfree rate. See main text for key to authors. | 121 |
| 3 | Definitions of the various forms of Proper Risk Aversion with the initial paper that developed the concept. The term “Very Weak Proper” is used here but Franke <i>et al.</i> do not give this property a name. Risk Vulnerability was called Weak Proper Risk Aversion in Gollier and Pratt’s 1993 working paper. . . . | 205 |
| 4 | Properties of the first four derivatives of the four utility functions used in this chapter at seven values in the range [1,4] with the coefficient of relative risk aversion equal to 3 at the point of next period’s expected consumption in each case. . . . | 227 |
| 5 | Properties of the first four derivatives of the four utility functions used in this chapter at seven values in the range [1,4] with the coefficient of relative risk aversion equal to 15 at the point of next period’s expected consumption in each case. . . . | 229 |

| | | |
|---|--|-----|
| 6 | Multi-period economy with certain and state dependent risky returns. The number of assets chosen by investors k_h^*, k_l^* if investors are fully informed of the investment policy in the high state and low state respectively. M is the metric $(k_l^* - k_h^*)/k_h^*$, quoted only for positive k_h^* . ΔU , also quoted only for positive k_h^* , is the change in expected utility for investors from concentrating investment uncertainty in the high state rather than the low state. For all simulations $s = 0.1, q = 0.4, \gamma = 3, y_h = 1, T = 100$. These results are based on 250 simulations in each case. | 280 |
| 7 | Definitions of data used in this study. Figures for raw data not seasonally adjusted unless stated (SA = Seasonally adjusted). The riskfree rate is the return on the 90 day treasury bill, market returns include dividend reinvestment. The Consumer Price Index for the US is for all urban consumers, RPI=Retail Price Index. ep=equity premium and rfr=riskfree rate. | 309 |
| 8 | Correlation coefficients for independent variables used in linear regressions for the US and the UK. | 312 |

- 9 An OLS regression of EP , EP^P , ΔR_f and ΔR_f^P against shocks in unemployment, inflation and changes in inflation. Figures in square parentheses associated with regression coefficients are t-statistics that have been Newey-West adjusted with Bartlett weights lag 12 or 7. The s.c. and het. diagnostics are LM test statistics (χ_1^2) for serial correlation and heteroscedasticity respectively, with associated p-statistics in square parentheses. Sample period: Jun 1971 – Mar 1993 (UK), Mar 1951 – Dec 1993 (US). The column headings should be read as follows: “ a, a^P, f, f^P ” is the constant in the regression, “ b, b^P, g, g^P ” is the coefficient of S_t , etcetera. 315

- 10 Average change in nominal riskless rate f_{ij} and cumulative abnormal equity premium p_{ij} around the time of an unemployment shock as modelled by S_t . Standard deviations (not errors) are given in parentheses. Sample period: Jun 1971 – Mar 1993 (UK), Mar 1951 – Dec 1993 (US) 321

- 11 The t-statistics associated with the Wilcoxon non-parametric test for between difference in means between “V Pos” and “V Neg” for pre and post event abnormal equity premiums and changes in riskfree rate. The t-statistic is quoted for the sample with the higher average values. 322

Acknowledgements

I would like to thank the members of the finance research community at Warwick Business School, particularly Jack Broyles, Martin Conyon, Ian Davidson, Stewart Hodges, David Oakes and Archie Pitts, for the encouragement that they have given to me and for all the proof reading that they have done. I am grateful to the participants of the following seminars run at Warwick Business School: the Financial Options Research Center (Chapter 3: February 1997, Chapter 5: November 1995, Chapter 6: January 1995), the Accounting and Finance Group (Chapter 4: February 1996) and the Doctoral Series in Finance (Chapter 4: May 1996). I would also like to thank two members of the Economics department at Warwick for their help with chapter 6: Jeremy Smith for his econometrics advice and Marcus Miller for his general comments. I would, of course, particularly like to thank my supervisor, Tony Steele, for all the assistance, encouragement and insights that he has given to me over the past six years.

Outside Warwick, I am particularly grateful to Sanjay Yadav of the Bank of England for his help with the econometrics in chapter 6. I would also like to thank Philippe Weil, Professor of Economics, Université Libre de Bruxelles and Co-director of the European Centre for Advanced Research in Economics and Christian Gollier, Institut d'Economie Industrielle, Université des Sciences Sociales de Toulouse for their useful comments on chapter 4. I would like to thank the editors of the European Journal of Finance for commissioning a version of chapter 3 for publication. I am grateful to the participants at the following seminar series: European Finance Association Doctoral Tutorial, (Chapter 5: August 1995, Milan), the British Accounting Association Doctoral Colloquium, (Chapter 6: April 1995, Bristol) and the Department of Accounting and Finance, Lancaster University (Chapter 6: June 1995).

I am particularly indebted to my family for their unstinting love and encouragement. My father has been a constant source of motivation and much sought after advice. My mother has helped me put work in perspective and kept the stress levels down. Both have given total and unconditional support to my education and career throughout my life. Without them, this thesis would probably never have been started let alone finished. I would also like to thank Katie, who has been the substantive original contribution to my own life during this period of research. And yes, brother, I do now know what the question is.

Part I

Introduction

1 Incomplete markets, financial puzzles

1.1 Issues to be addressed

“Because human capital represents such a large part of wealth, it is essential that future research focus on the effects of nontraded assets on individual behavior...”

Robert C. Merton (1992), p. 575

This dissertation examines whether theoretical models that explicitly incorporate sources of nontradable endowment risk are better able to explain observed financial market behaviour than models that assume that markets are complete (markets where all sources of financial risk can be fully insured). More specifically, this thesis examines whether two well documented financial market puzzles can be at least partially explained by relaxing assumptions of market completeness. The two market anomalies to be examined are the “Mehra & Prescott puzzles”¹ and the dividend controversy. The puzzles highlighted by Mehra and Prescott (1985) are that over the last century in the USA the average real riskfree rate has been lower, while the average excess return to the stock market over the riskfree rate has been higher, than standard complete market models can explain. The dividend controversy, discussed perhaps most famously by Black (1976), is that, despite the irrelevancy theorem of Miller and Modigliani (1961), dividend policy appears to continue to be a matter of concern for corporate treasurers and investors alike.

¹More usually referred to separately as the “riskfree rate puzzle” and “equity premium puzzle”.

That market incompleteness may play an important role in determining financial market behaviour has been recognised for some time. Leland (1968) introduced the concept of the “precautionary savings motive” — the idea that investors with convex marginal utility wish to save more *ceteris paribus* in the presence of consumption uncertainty than with a certain consumption stream. As introducing uninsurable risk increases the volatility of individual consumption, it is predicted that, to maintain equilibrium in financial markets, the riskfree rate is lower at times of high risk to personal capital than at times when such risk is low. The role of nonmarketable risk in a world with multiple risky assets has been under examination since early work by David Mayers. He showed that a version of the Capital Asset Pricing Model, where expected returns to marketable assets are linearly related to systematic risk (where systematic risk has a component measured against marketable income and a component measured against aggregate idiosyncratic income), remains valid in a mean-variance framework where not all assets are marketable both in the presence (Mayers (1972)) and absence (Mayers (1973)) of a riskfree asset. However, two fund portfolio separation is applicable in neither case as investors will choose a portfolio of marketable assets that will best hedge their idiosyncratic income. Mayers (1976) and Stapleton and Subrahmanyam (1979) developed conditions under which turning nonmarketable assets into marketable assets will (and will not) impact on the expected returns to the existing marketable assets in a mean-variance framework. However, these early models that incorporated nontradable risk did not appear to perform

well in practice. Fama and Schwert (1977) argued that Mayer's modified CAPM is almost observationally equivalent to the standard CAPM. This is because betas measured in a world with nontradable assets are virtually identical to betas measured against the market's return alone as covariances between aggregate human income and asset returns are very small.

More recently, the impact of nontradable risk in consumption based asset pricing models has come under consideration. It is within this paradigm that the current thesis is based. In particular, there is a growing literature that examines the role that uninsurable risk might play in helping to resolve the puzzles of Mehra & Prescott. Several authors, including Gregory Mankiw and Philippe Weil, believe that the existence of nonmarketable capital can help explain these puzzles as personal risks will significantly alter savings and consumption decisions. Others, notably John Heaton and Deborah Lucas, have argued that, even in incomplete markets, financial assets can be used to largely smooth nonmarketable risks. Therefore, in the absence of market imperfections, allowing for incompleteness does not alter the predictions of complete market asset pricing models in a way that significantly alters the predicted equity premium and real riskfree rate. The first major aim of this thesis is to contribute to this debate. To date, incomplete market models have not been applied to issues in corporate finance. The second aim of this thesis is to examine the dividend policy controversy in an economy where investors have nontradable risks. If aggregate investment is allowed to vary from the optimal level in a manner that investors cannot predict, it will be

shown that this combination of marketable and nonmarketable risks may provide a potential explanation for this puzzle.

The introduction proceeds as follows. Section 1.2 outlines the key characteristics of nontradable risks that will result in predictions that vary in a significant way from complete market models. Section 1.3 argues intuitively why such variations from the complete market model may be able to explain the anomalies under investigation. Section 1.4 describes the development of the dissertation and outlines the main areas of substantive original contribution.

1.2 Sources of nontradable risks.

The principal assumption that runs throughout this thesis and distinguishes this work from more traditional theoretical financial economics literature regards the nature of nontradable risk. Therefore it is useful to provide an overview of the economic assumptions regarding the form of the uninsurable risks that will drive the models that are to follow. This section sketches the main characteristics of the economies under review and provides intuitive justification for why such assumptions are necessary to get results that are significantly different from the complete market case. Such sources of nontradable risk are then justified from observations of real economies.

The central assumption underlying the results in this thesis is that there is a subset of individuals in the economy who will receive a stream of future income at least some part of which is currently uncertain. These future income streams are exogenous to the financial market models being constructed

so that this thesis is based in a partial equilibrium framework. Equally importantly, financial markets are incomplete so that people are unable to exchange their income stream for any other endowment path all elements of which are currently known. The income risk is said to be “uninsurable”, “nonmarketable” or “nontradable”. With this uncertain cash flow from personal capital, all individuals then decide how much to consume at present and how much will be invested for future consumption. Rates of return on financial assets are inferred through the assumption that equilibrium is maintained. The role of uninsurable income uncertainty in financial market behaviour can thus be determined.

The key difference between complete and incomplete markets is that “with complete markets, investors fully insure against idiosyncratic income shocks, and individual consumption is proportional to aggregate consumption. With limited insurance markets, however, individual consumption variability may exceed that of the average, and the implied asset prices may differ significantly from those predicted by a representative consumer model” (Heaton and Lucas (1993), p.1). The implications of complete market assumptions are well summed up by Cochrane (1991b) (his italics, pp. 957–9):

“If markets are complete ... then an individual’s consumption should not respond to idiosyncratic income or wealth shocks. This proposition can be viewed as a cross-sectional counterpart to the permanent income hypothesis: full insurance implies that consumption should not vary *across individuals* in response to idiosyncratic shocks, just as constant borrowing and lending opportunities imply that consumption should not vary *over time* in response to forecastable shocks ... Full insurance implies the existence of a representative consumer, that is, a social welfare function defined over aggregates that is independent of changes in the distribution of income or wealth over time. ”

If markets are complete financial economists can work in a representative agent world. However, as soon as complete market assumptions are dropped it is no longer clear that representative agent models (such as the Consumption CAPM) need follow. The economics literature contains many observations of consumption, wealth and portfolio composition that appear to violate the testable implications of complete frictionless market models:

“Casual empiricism as well as more formal evidence indicates that individual consumptions are much more volatile than aggregate consumption ... Individual wealth holdings appear to be highly volatile with large fractions of households moving from one wealth decile to another over a few years ... the ratio of median to mean income is higher for individuals in occupations with greater income uncertainty, e.g., farmers and self-employed businessmen ... The portfolios of households with low wealth contain a disproportionately large share of low return risk-free assets and a disproportionately small share of high return risky assets. The portfolios of high wealth households exhibit the opposite characteristic ... Last, it would be hard to reconcile the vast amount of trading in asset markets and the pattern of transaction velocities across assets with a complete frictionless market story. The above facts constitute quite strong a priori evidence in favor of the importance of uninsured idiosyncratic risk.”

Aiyagari (1994) pp. 662–3

This suggests that it may be fruitful to amend asset pricing models to account for the incompleteness of markets. In order to justify the introduction of uninsurable personal capital in financial economics, however, it is necessary to consider forms of nontradable risk that are both economically “real” and give significant deviations from the predictions of complete market models. Aggregate consumption data is observed to be highly smooth and as, for most investors, labour income is the single greatest source of wealth, the aggregate income process should also be modelled to have low variability. One

route is to model each individual's income stream as having low volatility. Existing literature (reviewed in the chapters to follow) will demonstrate that equilibrium asset prices in this case do not vary significantly (in an economic sense) from the complete markets case. An alternative approach is to assume that, at any point in time, a small number of investors will face large income uncertainty in future while the majority of the population will have no income uncertainty. This again will aggregate up to low aggregate consumption uncertainty but often leads to significant changes in consumption for a small number of investors. If, *ex-ante*, each member of the population does not know whether they will be in the high risk or low risk group, the savings behaviour of everyone might be expected to reflect the possibility of being in the high risk group. It might, therefore, be supposed that, in this case, financial market behaviour will be significantly different from the complete market case.

Can this type of nontradable risk be justified by observations of the types of uncertainty with which investors are faced in real economies? The concept of severe income risk that is uninsurable and only affects a small percentage of the population *ex-post*, but is a concern to most *ex-ante*, brings to mind unemployment². So, throughout the theoretical and descriptive sections of this dissertation (that is, excluding chapter 6, where the study looks explicitly at unemployment data), and in keeping with other work in this area, the term “unemployment” or “low probability, high impact shocks” will be used

²Several other risks in the real economy also take this form. Compulsory early retirement and long-term sickness are other obvious candidates.

to signify this type of nonmarketable income risk. So, not only is this type of risk likely to cause models to deviate from the complete market case, it is also possible to link this uncertainty with (at least one) type of risk faced by most investors in a real economy.

Other assumptions underlying the economies in this thesis are “standard”. Unless otherwise stated, these assumptions are as follows. There are no taxes or market frictions. All investors have homogeneous beliefs. Financial markets will consist of two assets (a riskfree asset and a “market” index). The riskfree asset will be assumed to be in zero net supply. There is only one consumption good which is instantaneously perishable. Investors share the same utility of consumption which is additively time-separable and has constant relative risk aversion — that is, utility is assumed to take power or logarithmic form for all investors.

1.3 Market anomalies and incompleteness

Preceding sections have briefly described the two financial market anomalies that are to be examined and outlined the form of nontradable risk that will cause incomplete market models to differ significantly from the complete market case. It has also been argued that the type of income uncertainty required to get significant deviations from complete market predictions can be likened to (at least) one source of risk in real economies — unemployment. This section briefly explains why these variations from the complete market case will help *explain*, as opposed to exacerbate, the anomalies under consideration.

With regard to the Mehra and Prescott puzzles, several authors have already examined the potential role that exogenous income risk might play in resolving these issues. The desire to precautionary save is seen as being a potential explanation for the riskfree rate puzzle. To also explain the puzzle of the equity premium using such endowment risk, the literature takes two divergent paths. First the property of proper risk aversion, which places restrictions on the first four derivatives of investors' utility functions, provides conditions under which introducing income risk will make an investor increasingly averse to *independent* marketable risk. Weil (1992a) uses such theory to partially explain the equity premium. This will be considered in detail in chapter 4. Alternatively, marketable risk and consumption risk can be modeled to be correlated. This approach was taken initially by Mankiw (1986), where personal endowment shocks are concentrated in periods with contemporaneous low dividends. Within this environment, the consumption beta of the market index is raised, thus increasing the predicted equity premium. Whether unemployment risk does, indeed, significantly alter financial market behaviour compared to the complete market case has been shown to depend on the persistency of the risk (see, in particular, work by John Heaton and Deborah Lucas that is reviewed below). With perfect financial markets, the consumption pattern of a long lived investor for whom periods of unemployment are short will be similar to the consumption pattern of the same investor with no unemployment risk. This is because, rather than saving in advance against the future risk of unemployment, the investor will

borrow and / or sell shares (including selling short, if necessary) at times of unemployment against positive future income shocks. Only with severe borrowing constraints or other market frictions will the savings/consumption decision of such an investor be influenced by short term unemployment risk. However, if periods of unemployment are long-term then investors must prepare in advance for the possibility of becoming unemployed. Therefore, in order to get equilibrium asset returns that vary significantly from the complete market case it is not only necessary to model individual income risk as resembling unemployment but it is necessary for periods of unemployment to be long lived or for there to be severe market frictions. Within this thesis, this is mainly achieved by creating one and two period models where investors, once unemployed, never become reemployed, which is in the style of both Mankiw (1986) and Weil (1992a).

No previous literature suggests that such uncertainty might also help explain the dividend controversy. It will be argued in this thesis that unemployment risk might, though, play a role in determining optimal aggregate dividends. Consider an economy with one risky firm that is all equity financed. Assume that real investment is under the control of the managers of this firm as opposed to the shareholders. That is, private individuals are restricted from investing directly in real assets (factories, machinery, etc.) and can only use the equity of the firm as a source for investment. It should, though, be emphasised that there is no conflict of interest between agents and principals in the model that is constructed. The key assumption is

that managers of the firm are not permitted to fully reveal their future investment plans (that is, investors are uncertain about future aggregate real investment) and, what is more, there is no guarantee that this investment will be at the optimal level at any time. So, within this model, the main source of marketable risk will come from not knowing future real investment rather than the (more usual) uncertainty regarding realised rates of return on capital. Let there also be periods when the risk to personal endowment is higher than at other time. There will be two effects at work. First, investors will want investment to be higher at times of low risk to personal capital as there is more money available for saving in this case. So positive investment shocks should be concentrated in states with low endowment risk. Second, it will be shown that, even if investment shocks have zero mean, they will still be more easily absorbed by investors when endowment is secure. That is, if investment uncertainty is concentrated in states with high income uncertainty then these effects combine to make investors more averse to the marketable risk. The implications of this are as follows. First, rights issues, which will coincide with periods of high investment, should be concentrated in bull markets when unemployment is low. An even stronger conclusion can be drawn: mean zero investment shocks should also be concentrated in high states. Dividends should be maintained in economic recessions as they have an important consumption smoothing effect. It is therefore contended that incomplete market models may prove a fruitful path for investigating financial management anomalies.

1.4 The development of this thesis

To finish the introduction, the development of the thesis is briefly described. This section also discusses the areas of substantive original contribution that are made by this dissertation.

Chapter 2 is a major literature review than considers the theoretical foundations upon which this thesis is based. First, the fundamental theorem of asset pricing, where a linear relationship exists between asset prices and the pricing kernel, is reviewed. In order to identify the pricing kernel, investor preferences are then briefly discussed. Throughout this thesis the baseline assumption is that investors have additively time separable power utility. This class of utility function is discussed and it is emphasised that the coefficient of relative risk aversion is the inverse of the elasticity of intertemporal substitution in this case. From this basis, the Euler equation is described. It is emphasised that representative agent applications of the Euler equation are generally only valid if markets are assumed to be complete. In incomplete markets, individual, as opposed to aggregate, consumption must be used in this equation. The CCAPM is then developed under various conditions in discrete time before the continuous-time proof of Breeden (1979) is presented. As will be highlighted in chapter 3, the Mehra & Prescott puzzles are just one of a number of anomalies that arise from consumption based asset pricing models and so it is very useful to show the links between the Euler equation and the CCAPM. The CCAPM also provides one of the four expressions for the predicted equity premium and real riskfree rate that will

form the basis for chapter 3. The chapter concludes by developing these four sets of equations.

Chapter 3 concentrates on a literature review of one particularly notable empirical violation of standard consumption based asset pricing theory — the Mehra & Prescott puzzles. While reviews of this area are now beginning to appear in the literature (see, for example, Kocherlakota (1996), Heaton and Lucas (1995), Siegel and Thaler (1997)), the author believes that this chapter makes substantive original contribution. The scope of the material covered makes this chapter arguably the most thorough review of the topic existent at the time of writing. Further, most of the prior studies concentrate on outlining potential explanations for the puzzles. Because of the foundations laid in chapter 2, the theoretical issues that are raised by these anomalies are clearly demonstrated here. The puzzles can be established from any of the four expressions for the real riskfree rate and equity premium given at the end of chapter 2. From this, it is not only clear what might help resolve them but also what can not. Potentially valid explanations are split into three categories: (i) that the *ex-post* realisations of asset returns in the US over the past century are not representative of the *ex-ante* expectations, (ii) that investor preferences are not well described by power utility with parameters close to current best estimates and (iii) that representative agent assumptions are unrealistic in markets both with and without frictions. Although no new results are presented, innovative demonstrations of existing ideas, particularly with regard to incomplete markets, helps clarify the links

between major studies. Chapter 3 concludes with a review of other empirical tests that have given results that are difficult to reconcile with standard consumption based asset pricing models. These tests not only provide estimates for the coefficients to place in the power utility function used later in the thesis but show the general difficulties that this paradigm has in explaining financial market behaviour. It is concluded that the Mehra & Prescott puzzles have been surprisingly difficult to resolve.

Chapter 4 considers in detail one paper that uses idiosyncratic risk to explain the Mehra & Prescott puzzles. In the model of Weil (1992a), marketable and income risks are independent. In this case, provided the investor is “Standard Risk Averse” (under the definition of Kimball (1993)), then the equity premium will be higher than in the complete market case. Chapter 4 has two distinct parts. Standard Risk Aversion is one of several varieties of “proper risk aversion”: a condition where combining two independent risks increases the aversion of an individual to one of them. The first section of chapter 4 develops an integrated approach to proper risk aversion for broad categories of small income gambles, including some that have not been considered by previous research. This integrated method also paves the way for the proof of a new result. It is shown for the first time that decreasing absolute risk aversion (DARA) and decreasing absolute prudence (DAP) will increase the risk aversion of an investor to a broader class of small income risk than has previously been recognised. The second section of the chapter considers the *quantitative* effects of proper risk aversion on equilibrium asset

prices. The predicted equity premium is calculated for Weil's example for four utility functions that are very similar locally (at the point of expected consumption) and that all exhibit DAP and DARA across the consumption domain. These utility functions vary from each other at very high level of derivatives (fourth or fifth derivative and above) and therefore are increasingly different from each other as consumption moves from its expected value. Given that unemployment is an "extreme" condition, although the utility functions are similar at the point of expected consumption, they are very different in the unemployment state. It is shown that the predicted equity premium is highly sensitive to the choice of utility functions. In particular it is shown that, even for Weil's example, DAP and DARA on their own are not sufficient to explain the *magnitude* of the observed equity premium. Given that existing literature finds it difficult to answer fairly basic questions about the form of investors preferences (are utility functions time separable? Are investors constant relative risk averse? Is 3 a realistic coefficient of relative risk aversion? ...) it is argued that the much more subtle issues regarding the form of utility function that are raised by Weil's example cannot be answered by existing results. This chapter concludes by suggesting future research that might help resolve the issue of determining the correct utility function to use in models that incorporate unemployment risk by estimating risk aversion at points of extreme consumption.

Chapter 5 applies some of the ideas of incomplete market theory to the corporate finance issue of dividend policy. In this case, marketable risk does

not come primarily from uncertainty over future profits. Instead, it is conjectured that, even at the portfolio level, investors cannot accurately predict the aggregate future investment plans of the underlying firms. Further, aggregate investment is not always at the optimal level. So, it is uncertainty over aggregate investment commitments that is the main source of marketable risk. It is shown that investors prefer periods of higher than expected investment to be concentrated in high states. Further, using simulations, it is demonstrated that even mean zero investment shocks are better absorbed by the market when there is a low risk to personal capital. It is argued that this is consistent with the observed financial market behaviour of dividend smoothing and rights issues concentrated in bull markets. Dividends play an important role at the aggregate level in helping investors to consumption smooth at a time when there is a high risk to idiosyncratic endowment.

Chapter 6 is the main empirical chapter. The chapter examines the model initially introduced by Mankiw (1986) who constrains marketable risk and unemployment risk to be highly correlated. This is done to increase the absolute magnitude of the covariance between the market's returns and the ratio of marginal utilities of consumption. As nonmarketable risk in this model is driven by unemployment, and taking the term "unemployment" literally in this context, this implies that unemployment surprises should be negatively correlated with stock market returns. Similarly, through the precautionary savings motive, the riskfree rate should drop (rise) prior to poor (good) unemployment news. This is an unusual characteristic as drops in

the riskfree rate would, *ceteris paribus*, be expected to be associated with a rise in the market index. This thesis provides estimates of the correlation between unemployment shocks and changes in asset returns. It is shown that stock market returns do indeed decline (rise) prior to “bad” (“good”) unemployment news in both the UK and the US; an observation that is consistent with the theoretical model of Mankiw (1986). However, the riskfree rate rises (falls) which is not consistent with the precautionary savings motive. A conclusion of this thesis is that it is difficult to reconcile the data with Mankiw (1986) style models. Chapter 7 concludes. It is clear that the testable implications of incomplete market models with nontradable risk that resembles unemployment risk are often very different from the testable implications of complete market models. That changes in consumption vary between individuals has been clearly established by the economics literature, which makes it hard to justify the use of complete market models. Further development of incomplete market models might well be expected to help explain financial market behaviour. This thesis suggests that even some puzzles in corporate finance might be usefully addressed within such a paradigm. However, current theories that incorporate nontradable risks do not appear to capture true behaviour. The rise in the riskfree rate prior to poor unemployment news is not consistent with precautionary saving as a potential explanation for the riskfree rate puzzle if unemployment is the main source of uninsurable risk. Further, the testable implications of many incomplete market models are highly sensitive to the form of the assumed utility function of investors.

The empirical evidence, while not conclusive, does suggest that, particularly within the US, individual income risk is not sufficiently persistent to explain Mehra & Prescott's puzzles. The conclusion of this thesis is that, while the *a priori* evidence suggests that nonmarketable risk should play an important role in determining equilibrium asset prices, current applications of incomplete market theory leave many issues unresolved.

Part II

Consumption based asset pricing

Consumption based asset pricing

Abstract

This chapter reviews the theoretical foundations upon which this thesis is based. The chapter starts with a derivation of the fundamental theorem of asset pricing and develops this into the Euler equation and Consumption CAPM. The continuous-time proof of the Consumption CAPM as derived by Breeden (1979) is then given. More recent theoretical developments in the area are then reviewed. The emphasis of the discussion is on examining how robust consumption based asset pricing models are to changes in underlying assumptions. The final section provides analytical forms for the equity premium and real risk-free rate in a complete market with no frictions and (time-separable) power utility. These equations will form the basis for discussing the Mehra & Prescott puzzles in the next chapter. It is concluded that the testable implications of consumption based asset pricing models are highly sensitive to assumptions of market completeness.

2 Introduction

This dissertation investigates whether the Mehra & Prescott and dividend puzzles can be at least partially explained by incorporating nontradable income shocks into the theoretical model of the economy. The role of this chapter is to review the literature on the theory of consumption based asset pricing which provides the foundation for this thesis. This theory is developed in more detailed discussion of the most relevant literature that can be found in other parts of the thesis. Chapter 3 looks in detail at the Mehra & Prescott puzzles and other empirical tests of consumption based asset pricing models. Chapter 3 also examines the role that market incompleteness might play in improving the explanatory power of these models. Chapter 4 examines the theory of proper risk aversion — the impact of income shocks on independent marketable risk — and chapter 5 briefly discusses the optimal dividend policy debate. The structure of this chapter is as follows:

- Section 3 develops the fundamental theorem of asset pricing. It is shown that if there is no potential for arbitrage then the price of any asset at time $t - 1$ is given by $p_{it-1} = E[(d_{it} + p_{it})\pi_t]$ where d_{it} is the dividend from the asset in the next time period and π_t is a “pricing kernel”. Subsection 3.2 reviews a one period, finite state approach to the theorem which portrays the pricing kernel in terms of the prices of pure securities. The extension of this result to an infinite state economy is given in subsection 3.3, where the proof is based on the mathematics of

Hilbert spaces. Appendix 9.4 demonstrates the strength and flexibility of the fundamental theorem. It is shown that the Capital Asset Pricing Model of Sharpe (1964), Lintner (1965), the Arbitrage Pricing Theorem of Ross (1976) and the options pricing model of Black and Scholes (1973) can all be developed directly from the fundamental theorem.

- In order to apply the fundamental theorem of asset pricing to consumption based asset pricing problems it is necessary to make assumptions about investor preferences in order to get explicit form for the pricing kernel. Section 4 describes the utility function most commonly applied to consumption based asset pricing problems: time separable power / logarithmic utility. This section concentrates on showing the link between the elasticity of intertemporal substitution and the coefficient of relative risk aversion for this form of preferences. We will return to this link in the next chapter.
- Section 5 combines the fundamental theorem of asset pricing with investor preferences in discrete time to give the Euler equation and the Consumption CAPM (CCAPM). One of the main difficulties that arises when applying consumption based asset pricing models is that the theory refers to the consumption of individual investors. It is extremely difficult to work in these terms and so it is desirable to replace individual consumption with aggregate consumption in these models. Subsections 5.1 and 5.2 provide aggregation conditions that enable us to work

in a representative agent environment. The key assumption for this is that markets are complete. Therefore, in the remainder of this thesis, where it is assumed that there is uninsurable risk, it is the use of aggregate, as opposed to individual, consumption within consumption based asset pricing models that is the point of contention. Having provided conditions for aggregation, subsections 5.3 and 5.4 then derive the Euler equation from the fundamental theorem of asset pricing both with and without a representative agent. The Euler equation is the central model that underlies the work in this thesis. In subsections 5.5 and 5.6 the link between the Euler equation and the discrete time CCAPM is shown. Essentially the CCAPM can be derived from the Euler equation provided that we can aggregate, use a “sensible” utility function and provided that aggregate consumption is sufficiently smooth. This link between the CCAPM and Euler equations is important as it shows that the Mehra & Prescott puzzles, which are usually demonstrated via the Euler equation, are essentially a CCAPM puzzle. The Mehra & Prescott puzzles should therefore be interpreted in the light of other empirical tests of the CCAPM. This is discussed further in the next chapter.

- Section 6 develops the CCAPM in a continuous-time environment. Subsection 6.1 gives the Intertemporal CAPM of Merton (1973). Following Breeden (1979), this is simplified into the CCAPM in section 6.2. In subsection 6.3 the literature that has relaxed some of the underlying

assumptions of the CCAPM is reviewed. In particular the work of Back (1991) and Aase (1993) is reviewed, where there are jumps in the optimal consumption path of each individual and asset returns. Subsection 6.4 contrasts the CAPM with CCAPM. It considers economies where the two models are equivalent and presents a model of equilibrium asset prices where systematic risk has both a market and consumption component.

- Section 7 uses the theory developed in the earlier sections to derive analytical forms for the equity premium and real riskfree rate. This is done in non-parametric single period, parametric single period, continuous time and multi-period discrete time Markov growth environments. It will be shown in the next chapter that the puzzles of Mehra & Prescott can be demonstrated using any of these models. What becomes clear is the assumptions that drive the puzzles: (i) a representative agent exists (ii) there are no market frictions or taxes (iii) investors have additively time-separable power / logarithmic utility with parameters that appear to reflect investor preferences and (iv) aggregate consumption is smooth.

The *contribution* of this chapter is as follows. While no new results are presented, this is, to the author's knowledge, the most comprehensive review of the theory underlying the Mehra & Prescott puzzles. Reviews of the puzzles (Heaton and Lucas (1995), Kocherlakota (1996), Siegel and Thaler (1997)) have concentrate more on providing potential explanations than ex-

plaining the origins of the puzzles. Original papers have chosen one theoretical basis for developing the puzzle (*e.g.* Mehra and Prescott (1985) use a Markov growth model, Ahn (1990) uses a dynamic programming approach). To the author’s knowledge, this is the only source that provides four related, and yet separate, formulations for the real riskfree rate and equity premium. The author believes that, in providing such a solid theoretical foundation for the puzzles, the role of each of the individual assumptions becomes clear. Perhaps equally importantly it is clear what can *not* explain away the puzzles — for example discrete time / continuous time arguments or subtle debates about the exact process describing aggregate consumption³ or asset returns. In particular, the role that market completeness plays in models of asset prices is clear and the papers presented in subsection 6.3 hint at how the Mehra & Prescott puzzles might be resolved by allowing for uninsurable risks.

3 Fundamental theorem of asset pricing

3.1 Introduction

The aim of this section is to provide a very general theoretical setting for the work that is to follow. Under one of the most widely accepted assumptions regarding investor preferences — that there exists an investor who is never satiated — it is possible to deduce a highly generalised theorem of asset pricing. This result, known as the “fundamental theorem of asset pricing”,

³Although the debate about potential jumps in aggregate consumption might prove crucial as this violates the central assumption of smooth aggregate consumption.

states that there must exist a positive linear operator that will value any future set of risky cash flows. The theorem was developed in a one period, finite state economy by Stephen Ross (Ross (1977), Ross (1978)) and developed by Harrison and Kreps (1979) into a multiperiod, infinite state economy and Harrison and Pliska (1981) into a continuous-time economy⁴. Individual asset pricing models, such as the CAPM and Consumption CAPM (CCAPM), can be linked through this theorem with the specific linear operator varying from model to model. This section of the thesis develops the fundamental theorem of asset pricing so that later sections can place consumption based asset pricing models within this setting.

The positive linear operator is presented in different ways by different authors:

“There are many equivalent ways of representing a linear pricing rule ... In one representation, the price is the expected value under artificial ‘risk neutral’ probabilities discounted at the riskless rate. (The risk-neutral probability measure is also referred to as an equivalent martingale measure). In another representation, the price is the expectation of the quantity-times-state-price density, which is the state price per unit probability. In yet another representation, the price is the expected value discounted at a risk-adjusted rate.”

Dybvig and Ross (1992) pp. 46–7

The section proceeds as follows. First, in order to provide the clearest interpretation for the pricing kernel, the fundamental theorem of asset pricing is presented in a one period, finite state, economy. The theorem is then

⁴There are several text book treatments of the area. For example, Ingersoll (1987), Huang and Litzenberger (1988) and Ferson (1995) provide accessible accounts while Darrell Duffie (Duffie (1988), Duffie (1992a)) provides more rigorous vector space treatment of the type given here. Dybvig and Ross (1992) provide an excellent synopsis of the topic while Constantinides (1989) provides an introductory overview of the interrelations between the various asset pricing models.

developed in subsection 3.3 following Harrison and Kreps (1979). The proof relies on the mathematics of Hilbert spaces. Appendices are provided on the relevant pure mathematical background.

3.2 The fundamental theorem in economies with finite states

In this subsection, the fundamental theorem of asset pricing is developed in a single period, finite state economy. This subsection is included because it provides clear interpretation for the pricing kernel in terms of the price of pure securities. As the pricing kernel will be central to the development of this thesis, the author considers it important to present an intuitive representation for this variable.

Before proceeding, the following remark helps place any arbitrage-free⁵ pricing theorem within an equilibrium⁶ environment:

Remark 1 *A market that presents arbitrage opportunities cannot be in equilibrium provided there exists at least one investor who is never satiated.*

⁵As observed by Ingersoll (1987) (p.52 *et. sec.*), there are two types of arbitrage opportunity. An arbitrage opportunity of the first type is a portfolio with non-positive cost and payouts that are never negative and have non-zero probability of being positive. An arbitrage opportunity of the second type is a portfolio with negative cost and payouts that are strictly non-negative. The potential for arbitrage of the first type neither implies, nor is implied by, an arbitrage opportunity of the second type. Strictly speaking, this section provides proofs only for arbitrage of the first type. However, the results hold equally for arbitrage opportunities of the second type and the distinction need not concern us here.

⁶Formal definitions of competitive market equilibrium will follow when the theoretical models are being developed below. Intuitively equilibrium in a financial market refers to an allocation of assets amongst investors and a set of prices that is both feasible (the sum of the holdings in each asset is equal to the total supply of that asset) and such that, for the prevailing market prices, the portfolio of assets owned by each individual maximises the expected utility of that individual subject to the individual's budget constraint. At the simplest level, a financial market is in equilibrium if there are no individuals who wish to change portfolio at the given market prices. See, for example, p.1 of Duffie (1988) for a simple mathematical description of an equilibrium market.

If arbitrage opportunities exist, there is the opportunity for any investor to create a *free lunch* by investing in this opportunity. If there exists a subset of investors who are never satiated, these individuals should wish to take an infinite position in this free lunch. This is clearly not consistent with equilibrium.

The fundamental theorem of asset pricing is now presented for a single period, finite state economy. Consider a single period economy with S (finite) potential future states (in an Arrow-Debreu sense) at time $t = 1$. Let there also be n (finite) assets in the world. Let the $(n \times S)$ payoff matrix at $t = 1$ be denoted by \mathbf{D} and let $\mathbf{D}_i, D_{i,s}$ denote the i^{th} row of this matrix and the element in column s of this row respectively. Using \mathbf{R}_{++}^S to denote a $(S \times 1)$ vector of strictly positive real numbers and \mathbf{p} as the $(n \times 1)$ vector of prices of the n assets at time $t = 0$, the only time when trading is permitted⁷:

Result 1 *The economy permits no arbitrage if and only if $\exists \varpi \in \mathbf{R}_{++}^S$ such that $\mathbf{p} = \mathbf{D}\varpi$.*

In the course of demonstrating the fundamental theorem, a second important result will also be shown to hold:

Result 2 *Suppose that the market permits no arbitrage. Then $\varpi \in \mathbf{R}_{++}^S$ that satisfies $\mathbf{p} = \mathbf{D}\varpi$ is unique if and only if the market is complete.*

The proof of this result is presented in three steps. First it will be shown that the assumption that $\exists \varpi \in \mathbf{R}_{++}^S$ such that $\mathbf{p} = \mathbf{D}\varpi$ cannot hold if the

⁷Throughout this thesis the standard mathematical notation \exists, \forall are used to denote “There exists...” and “For all...” respectively.

economy permits arbitrage. Second it will be argued that such a ϖ must exist in a complete arbitrage-free market and that ϖ will be uniquely identified in this case. Third, it will be argued that if the market is incomplete then ϖ will not be uniquely defined.

First suppose that the economy does permit arbitrage. In this case it will be shown that no ϖ of the form required by the result can exist. By the definition of arbitrage, it is possible in this economy to come up with a self-financing trading strategy that will not give a negative payout in any state and will give a positive payout in at least one state. Denote the holdings in the n assets that provide this arbitrage opportunity by an $(n \times 1)$ vector θ so that the payout in the S states is given by $D^T \theta$. The cost of this self-financing portfolio is $p^T \theta \leq 0$. Suppose that $\exists \varpi$ such that $p = D\varpi$. By the assumption of self financing, $p^T \theta \leq 0 \Rightarrow \varpi^T D^T \theta \leq 0$. By the assumption of arbitrage $D^T \theta$ is never negative and positive at least once. Then $\varpi_i \leq 0$ for at least one i . This contradicts the assumption that $\varpi \in \mathbb{R}_{++}^S$. Next suppose that the market is complete and there are no arbitrage opportunities. Then, by definition of completeness in an Arrow-Debreu economy, it is possible to create the S pure securities⁸. It may be possible to create the pure securities using more than one combination of the n assets, but, however we create these securities, the absence of arbitrage ensures that the price of the pure securities are uniquely defined. We can therefore denote the prices of the S pure securities by the unique $(S \times 1)$ vector ϖ . Given the absence of arbitrage

⁸The pure security for state s is an asset that pays 1 if the state of the world is s and 0 otherwise.

and the fact that the pure securities give a positive payout in one state and never a negative payout, $\varpi \in \mathbb{R}_{++}^S$. The payouts from any asset i can be replicated by buying $D_{i,s}$ shares in all $s \in [1, S]$ pure securities. The absence of arbitrage ensures that $\mathbf{p} = \mathbf{D}\varpi$ as required. This gives an economic interpretation to the vector ϖ . ϖ is the price of the S pure securities and will be referred to as a “state-price deflator”.

Given this interpretation of ϖ , the situation clearly becomes more difficult in the case of incomplete markets and no arbitrage. An incomplete market without arbitrage can be considered to be an arbitrage-free complete market with certain assets withdrawn. That is, by adding in somewhere between 1 and S additional assets, it will be possible to complete the market without introducing arbitrage opportunity. In this enlarged market, it has already been shown that the state-price deflator exists. Therefore, it must also exist in the restricted incomplete market. It has already been shown that in a complete market ϖ is unique (as it is the price of the pure securities). Therefore, in the case of an incomplete market, ϖ is going to be unique if and only if, for any extension of the market that makes it complete and arbitrage free, the price of all pure securities will be the same. This will never be the case since adding new assets introduces important degrees of freedom. This is demonstrated by example. Consider the case of $n = 2, S = 3$ with prices and payout matrix given below. This market is arbitrage-free and so we know that $\mathbf{p} = \mathbf{D}\varpi$:

$$\begin{pmatrix} 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 2 & 6 & 8 \\ 2 & 8 & 9 \end{pmatrix} \begin{pmatrix} \varpi_1 \\ \varpi_2 \\ \varpi_3 \end{pmatrix} \quad (1)$$

It can be shown in this case that ϖ is not unique by presenting two admissible values for this variable. The market could be completed by adding a state 1 pure security with price 0.8 or 0.9. Solving: $\varpi = (0.8, 0.46, 0.08)^T$, $(0.9, 0.48, 0.04)^T$ respectively. Both these vectors are in \mathbf{R}_{++}^3 and so the market is still arbitrage free in both cases. From this, it is clear that there is a continuum of $\varpi \in \mathbf{R}_{++}^3$ that satisfy equation 1. So, for an incomplete market, the state-price deflator is not unique, but at least one ϖ of the form required by the theorem does exist. The result has thus been established.

As emphasised in the quotation of Dybvig and Ross (1992), several interpretations can be placed on the linear relationship between prices and payouts. The theory has been developed so far in terms of the state-price deflator. Two alternative representations are now presented — the existence of a “pricing kernel” and of an “equivalent martingale measure” (EMM)⁹. Use P_s to denote the probability that all investors ascribe at $t = 0$ to the economy ending in state s at $t = 1$ ¹⁰ and $E^x[\cdot]$ to denote expectations under probability measure x . Consider table 1. Q is, indeed, a probability measure. All values are non-negative as $\varpi \in \mathbf{R}_{++}^S$ and it clearly sums to one.

⁹A formal definition of an equivalent martingale measure is given in appendix 9.1.

¹⁰Throughout this thesis it is assumed that all investors agree on the probability space (see, again, appendix 9.1). That is, throughout this thesis “beliefs” about the future state of the world are homogeneous across the investment community. This thesis does not consider the role that different information sets and different expectations across the community of investors might play in determining equilibrium asset prices in financial markets.

$Q_s = 0$ if and only if $\varpi_s = 0$, which will only happen if the probability of state s is zero under the initial probability measure P . Further, $\sum_{s=1}^S \varpi_s$ has interpretation. In a complete market it is the cost of creating a portfolio of all the pure securities: that is, it is the cost of the riskless portfolio. Therefore $\sum_{s=1}^S \varpi_s = 1/(1 + r_f)$, where r_f denotes the riskfree rate. So all assets payoffs, discounted at the riskfree rate, are martingale under the new probability measure Q . This makes Q a well defined equivalent martingale measure.

Of more direct relevance to this thesis is the representation of the linear pricing rule as the strictly positive $(S \times 1)$ vector π which will be called the “pricing kernel”. We know that each element π_s in π is strictly positive as both ϖ_s and P_s are strictly positive for all $s \in [1, S]$. That, given the absence of arbitrage opportunities, the price of any asset i at time 0 can be given as $E[d_i; \pi]$ (where d_i is the total payout to the asset at $t = 1$, π is some strictly positive pricing kernel and expectations are taken with respect to the probability space agreed on by all investors in the community) will prove a valuable foundation upon which to develop the theoretical discussions in subsequent sections.

3.3 A Hilbert space proof of the fundamental theorem

This dissertation is not restricted to finite state economies. The proof of the fundamental theorem in an infinite state, single period economy is now

| Pricing kernel | EMM |
|---------------------------------------|---|
| $\pi_s := \frac{\varpi_s}{P_s}$ | $Q_z := \frac{\varpi_z}{\sum_{s=1}^S \varpi_s}$ |
| $p_i = \sum_{s=1}^S D_{is} \pi_s P_s$ | $p_i = \sum_{s=1}^S \varpi_s \sum_{z=1}^S D_{iz} Q_z$ |
| $= E^P[D_i \pi]$ | $= \sum_{s=1}^S \varpi_s E^Q[D_i]$ |

Table 1: Representing the linear relationship between prices and payouts as a positive pricing kernel and equivalent martingale measure (EMM).

presented. This comes from the work of Harrison and Kreps (1979)¹¹. This section calls on some pure mathematics on vector spaces that will not be explicitly used in the rest of the thesis. A brief discussion of these issues can be found in appendix 9.2 and references are given in the body of the text to more detailed treatments.

The notation is as follows. Consider the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ (see appendix 9.1) and, for any continuous linear (\mathcal{F} measurable) functional x from Ω to the real line, define the norm $\|x\|$ by¹²:

$$\|x\| = \left(\int_{\Omega} |x(\omega)|^2 \mathcal{P}(d\omega) \right)^{1/2}$$

Define $X \equiv L^2(\mathcal{P}) := \{x \in L : \|x\| < \infty\}$. These square integrable func-

¹¹The proof in the original paper is for an economy consisting of T (finite) trading periods where no funds are either introduced or removed from the market over these time periods. This subsection deals only with single period economies.

¹²Technical note: If two measurable functions x, y are equal almost everywhere then $\|x - y\| = 0$. $x = y$ almost surely but $x - y \neq 0$. Therefore, strictly speaking, this is not a well defined norm. To avoid this problem it is assumed here that no two linear functionals are equal almost everywhere and leave a more formal treatment of this problem to the textbooks.

tionals can be considered to be possible payouts on financial assets. Consider the subset $M \subseteq X$ that represents the payouts to traded assets. $X \setminus M$ is then the set of potential contingent claims. The proof of the fundamental theorem comes from the observation that X is a Hilbert space:

Result 3 *$L^2(\mathcal{P})$ is a Hilbert space with a norm defined above and an inner product defined by $(x|y) = E^{\mathcal{P}}[xy]$*

Borowski and Borwein (1989) refers to L^2 spaces as “the most common realisations (of Hilbert spaces)” (p.267)¹³. Having established that $L^2(\mathcal{P})$ is a Hilbert space, the Rietz Representation Theorem for Hilbert spaces is now presented. See, for example, p.137 of Bollobás (1990) or page 109 of Luenberger (1969) (where this result is called the Riesz-Fréchet theorem) for a full derivation, but an outline of the proof is given in appendix 9.3:

Result 4 *Let f be a continuous linear functional on the Hilbert space H . Then there is a unique element $y \in H$ such that $f(x) = (x|y)$ for all $x \in H$*

The following corollary is of particular importance, which can also be found in Royden (1968), p.246:

Corollary 1 *Let p be a continuous linear functional on $L^2(\mathcal{P})$. Then there is a unique element $\pi \in L^2(\mathcal{P})$ such that $p(x) = E^{\mathcal{P}}[x\pi]$ for all $x \in L^2(\mathcal{P})$*

The existence of a unique, strictly positive, pricing kernel with finite variance in a complete market¹⁴ follows directly as, in the absence of arbitrage,

¹³The proof of this result is not given here as it is so widely available in the literature — see, for example, p.134 of Bollobás (1990) or p.210 of Royden (1968).

¹⁴In this context a complete market is a market where either $M = X$ or the price of all contingent claims in $X \setminus M$ are priced by arbitrage by the set of assets in M .

the pricing function p will be a continuous linear functional and π must be strictly positive given that p is a strictly increasing functional.

That π is not unique in an incomplete market is not formally shown here — see the original paper. It is clear, though, from the discussion given above: an incomplete market can be made complete in many different ways without introducing arbitrage opportunities. Each of these complete markets will have a different pricing kernel and so the pricing kernel is not unique in an incomplete market. The formal development of pricing kernels in a T period (T may be finite or infinite) discrete time economy is also not given here. However, by repeated substitution into the fundamental theorem, it is “clear” that, for any asset i that pays dividend $d_{i\tau}$ at time τ , the price of the asset at time t , p_{it} is given by:

$$p_{it} = \sum_{\tau=t+1}^T E_t[d_{i\tau}\pi_{\tau}]$$

for some strictly positive π_{τ} which is independent of i . There are five other versions of this relationship which will be useful¹⁵:

$$1 = E[(1 + r_{it})\pi_t] \quad (2)$$

$$0 = E[(r_{it} - r_{jt})\pi_t] \quad (3)$$

$$E[r_{it}] = r_{ft} + \frac{\text{Cov}(r_{it}, -\pi_t)}{E[\pi_t]} \quad (4)$$

$$E[r_{it}] = r_{ft} + E[r_{jt} - r_{ft}] \frac{\text{Cov}(r_{it}, -\pi_t)}{\text{Cov}(r_{jt}, -\pi_t)} \quad (5)$$

¹⁵Notice that, while equation 4 implies equation 5, the implication does not run the other way. Therefore these equations are not equivalent.

$$r_{ft} = \frac{1}{E[\pi_t]} - 1 \quad (6)$$

for any assets i, j and the riskless asset whose return is denoted by r_{ft} and where $r_{it} := (p_{it} + d_{it} - p_{it-1})/p_{it-1}$ is the simple (discrete) return to asset i . The secret of asset pricing is now to identify π_t . The strength of the fundamental theorem is that *all* asset pricing relations can be interpreted in terms of a pricing kernel. To demonstrate the strength and flexibility of the theorem, the CAPM, APT and Black-Scholes options pricing formula are all derived from the fundamental theorem in appendix 9.4. This thesis is, though, concerned with consumption based asset pricing models. In section 4 investor preferences are discussed which will enable π_t to be interpreted as the ratio of marginal utilities in the Euler equation (subsections 5.3, 5.4). This will form the basis for the consumption based asset pricing models that lie at the heart of this thesis.

4 Investor preferences

In the previous section the fundamental theorem of asset pricing was discussed. Before deriving consumption based asset pricing models and placing them within the context of the fundamental theorem, this chapter must discuss investor preferences. This is necessary as the pricing kernel will be identified as the ratio of marginal utilities in the Euler equation. Different assumptions are made by different authors and these will be highlighted in the discussions to follow. This section aims to outline the “most common” assumptions about preferences that underlie these models.

In general, it is assumed that there is only one consumption good and that utility is additively time-separable and state independent. That is (in discrete time with an obvious continuous-time analogy), if an investor is going to consume c_1, c_2, \dots, c_T amounts of the consumption good at times $t = 1, 2, \dots, T$ during her remaining lifetime before known time of death T , then she derives utility $u(c_1, c_2, \dots, c_T) = U(c_1, 1) + U(c_2, 2) + \dots + U(c_T, T)$. Here $u(c_t, t)$ denotes the utility that results from consuming a quantity c_t at time t ¹⁶. The next key assumption regards the form of $U(c_t, t)$. It is usually assumed that $U(c_t, t) = \beta^t U(c_t)$ where $U(c_t)$ is now independent of t and β reflects the time preference of investors. It is generally believed that investors prefer to consume sooner rather than later (*ceteris paribus*) so that $\beta < 1$. Therefore the form of the utility function is:

$$u(c_1, c_2, \dots, c_T) = \sum_{t=1}^T \beta^t U(c_t)$$

With this time-separable utility it is now necessary to make assumptions about the form of $U(c_t)$. The usual assumption is that the utility function takes the following form:

$$U(c_t) = \begin{cases} \frac{c_t^{1-\gamma} - 1}{1-\gamma} & \gamma \neq 1 \\ \ln(c_t) & \gamma = 1 \end{cases} \quad (7)$$

Clearly the power form is not well defined for $\gamma = 1$ but, by l'Hôpital's

¹⁶A utility of bequest function $B[W_T, T]$ is usually added to this to account for the utility that the investor gets from leaving wealth W_T to the next generation at time T (assuming that the investor is finitely lived). This section ignores the bequest function for algebraic simplicity only.

rule, the limit of the power form as γ tends to one is the logarithmic form¹⁷. This utility function will be given three names interchangeably during the remainder of this thesis. It will be called “power utility” (with logarithmic utility implicitly assumed as a special case), “isoelastic utility” or “constant relative risk aversion (CRRA) utility”¹⁸. In some cases, power utility will be arbitrarily rescaled to $U(c) = c^{1-\gamma}/(1-\gamma)$ with no loss of generality for algebraic simplicity. These utility functions have the important property that, under “standard” assumptions, investors will invest a constant proportion of their wealth in risky assets. This contrasts with constant absolute risk aversion utility (exponential utility) where investors will invest a constant amount of money in risky assets (see, for example, p.118 of Merton (1992) for this comparison). The concept that someone worth £10,000 invests the same quantity in the stockmarket as someone worth £10 million seems unrealistic and argues against exponential utility. The concept that both these investors might invest 30% (say) of their wealth in the stockmarket is more reasonable and is loose support for CRRA utility functions. Early empirical support for this was given by Friend and Blume (1975): “Perhaps the most accurate single statement is: if there is any tendency for increasing or decreasing proportional risk aversion, the tendency is so slight that for many

¹⁷l’Hôpital’s rule (see, for example, p.104 of Binmore (1977)) states that if $f(a) = g(a) = 0$ then $\lim_{x \rightarrow a} f(x)/g(x) = \lim_{x \rightarrow a} f'(x)/g'(x)$. Substituting in $f(x) = c^{1-x} - 1$, $f'(x) = d(c^{1-x})/dx = d[\text{Exp}[\ln(c^{1-x})]]/dx = d[\text{Exp}[(1-x)\ln(c)]]/dx = -\ln(c)\text{Exp}[(1-x)\ln(c)]$. So, $\lim_{x \rightarrow 1} f(x) = -\ln(c)$. If $g(x) = 1 - x$ then $\lim_{x \rightarrow 1} g'(x) = -1$.

¹⁸The last two names are somewhat loose. Some time non-separable utility functions have constant relative risk aversion and some have constant coefficients of intertemporal elasticity of substitution. Unless specifically stated to the contrary, utility functions will be assumed to be additively time separable in the remainder of the thesis.

purposes the assumption of constant proportional risk aversion is not a bad first approximation" (*ibid.* p. 915)¹⁹. This reason, together with its tractability, explains the popularity of power utility functions for consumption based asset pricing²⁰.

Before proceeding, there are two terms related to utility that are necessary for the discussion that is going to follow. Consider a utility function that has two parameters x, y , so utility is $u(x, y)$. The *marginal rate of substitution*, MRS_{yx} , of y for x at a point of consumption \bar{x}, \bar{y} is defined by $u(\bar{x} - 1, \bar{y} + MRS_{yx}) - u(\bar{x}, \bar{y}) = 0$. Assuming without loss of generality (through arbitrary rescaling) that x, y are large relative to 1, MRS_{yx} , a first order Taylor's series expansion gives:

$$MRS_{yx} = \left(\frac{\partial u / \partial x}{\partial u / \partial y} \right) \Big|_{x=\bar{x}, y=\bar{y}}$$

Notice that, in equilibrium, MRS_{yx} can be considered to be the price of consumption good x in terms of consumption good y . The *intertemporal marginal rate of substitution*, $IMRS_{\tau t}$, between time τ and t is the marginal rate of substitution of consumption at time τ for consumption at time t . That is, the IMRS is the ratio of marginal utility of consumption at t to marginal utility of consumption at time τ . However, as consumption at time t, τ may not be known with certainty at the time when the IMRS is being

¹⁹If anything their evidence supports increasing rather than decreasing relative risk aversion.

²⁰Although exponential utility is sometimes used. See the next chapter where exponential utility will provide closed solutions to certain precautionary savings models that are intractable under power utility. See Stapleton and Subrahmanyam (1978) for an equilibrium asset pricing model based on exponential utility.

calculated, it will be useful in general to think of the IMRS as stochastic.

$$\text{IMRS}_{\tau t} = \left(\frac{\partial u / \partial c_t}{\partial u / \partial c_\tau} \right) \Big|_{c_t = \bar{c}_t, c_\tau = \bar{c}_\tau}$$

The $\text{IMRS}_{t\tau}$ can be interpreted as being the price of consumption at time τ in terms of consumption at time t . The IMRS then has an obvious interpretation as being related in some way to the interest rate. This point will be returned to in the next section on the Euler equation. The *elasticity of substitution* between two goods x, y , ES, is a measure of how the ratio of the level of consumption of x, y varies as the ratio of prices of x, y vary (see, for example, p.198 et. sec. of Green (1976) for a relevant discussion). Given that, in equilibrium, prices are given by marginal rates of substitution:

$$\text{ES} = \frac{\Delta(x/y)}{x/y} \frac{\text{MRS}_{xy}}{\Delta(\text{MRS}_{xy})}$$

The *elasticity of intertemporal substitution*, EIS, between time t, τ is the elasticity of substitution of consumption between time t and τ . That is,

$$\text{EIS} = \frac{\Delta(c_\tau/c_t)}{c_\tau/c_t} \frac{\text{IMRS}_{\tau t}}{\Delta(\text{IMRS}_{\tau t})}$$

It will now be shown that the utility function given in equation 7 has constant EIS with the elasticity of intertemporal substitution being $1/\gamma$ at all points. This is demonstrated by considering the utility functions that do have $\text{EIS} = 1/\gamma$ at all points and then showing that the functions in equation 7 are members of this class. So, consider the utility functions for which $1/\gamma$ is the EIS at all points. For notational simplicity let $X := c_\tau/c_t$ and

$Y := \text{IMRS}_{\tau t}$. Taking the previous definition of EIS for infinitesimal changes, this means that $dY/Y\gamma = dX/X$. Integrating out, $\ln(Y)/\gamma = k + \ln(X)$ for constant of integration k . So, $Y^{1/\gamma} = KX$ where $K := e^k$. Substitute back in for original notation:

$$\text{IMRS}_{\tau t}^{1/\gamma} = \left(\frac{\partial u / \partial c_t}{\partial u / \partial c_\tau} \right)^{1/\gamma} = K(c_\tau / c_t) \quad (8)$$

By substituting equation 7 into equation 8 it is now clear that the elasticity of intertemporal substitution is $1/\gamma$ at all points for power utility functions.

Next it will be shown that the utility function given in equation 7 has constant relative risk aversion in the sense of Pratt (1964) and Arrow (1970). The coefficient of relative risk aversion $\mathcal{R} := -U''(c_t)c_t/U'(c_t)$. By simple substitution of equation 7 into this definition, it is clear that $\mathcal{R} = \gamma$ at all points for power utility functions.

That γ has two distinct roles within this utility function is of central importance in some of the discussions that follow — particularly subsection 12.2. It provides information about how sensitive an investor's savings / consumption decisions are to changes in the underlying interest rate. Second, it provides information on the investors' aversion to instantaneous gambles.

While these two characteristics may be economically related²¹ there is no

²¹If an investor is highly averse to an instantaneous gamble then γ will be high. It might also be reasonable to suppose that this investor would then also have a high desire to smooth consumption across time as well as across states. The investor would then have a low elasticity of intertemporal substitution — that is, the savings / consumption decision is dominated by the desire to smooth consumption rather than being driven by the incentive to save. The low EIS implies $1/\gamma$ is low, or γ is high. So linking risk aversion

reason necessarily why this should be so. Turning the EIS and \mathcal{R} into separate degrees of freedom within the utility function will be an important factor in alternate utility functions that will be considered in the next chapter.

5 The Euler equation and discrete time theory

This section derives the stochastic Euler equation, which was developed by LeRoy (1973), Rubinstein (1976) and Lucas (1978). The Euler equation will hold for all agents under widely varying assumptions about the nature of the underlying economies. This section concentrates on developing sufficient rather than necessary conditions. Following Lucas (1978), the discussion given here is for pure exchange economies and concentrates on one period models. Asset prices reflect the saving and consumption decisions of all agents who have access to the market. Sometimes, though, prices behave *as if* there were one “representative agent” in the economy who receives the aggregate endowment. The first section will show that, in a complete market, the ratio of the marginal utility of consumption at $t = 1$ to the marginal utility of consumption at $t = 0$ will be fixed across investors whatever state occurs at $t = 1$. So, an implication of complete markets is that any idiosyncratic shock in an individual’s endowment at $t = 1$ should not influence the consumption of that individual at $t = 1$ as, with full insurance available, the

with elasticity of intertemporal substitution in the manner of isoelastic utility can be economically interpreted by observing that if an investor has a high desire to consumption smooth across states instantaneously then she will also have a high desire to consumption smooth across time.

individual will have fully insured all personal endowment risks away at $t = 0$. This provides the basis for the next section, which provides a sufficient condition for the existence of a representative agent with power utility. These conditions are that the market is complete and that each agent has power utility with identical coefficients of time-preference and relative risk aversion. Next, the Euler equation is established in an economy with a representative agent — this is the proof of Lucas (1978). Finally, it will be argued that the Euler equation must also hold for each individual investor in an incomplete market where the representative agent need not exist.

5.1 Complete markets and aggregate consumption

This section aims to show that the consumption pattern for each individual is influenced only by aggregate consumption and not the allocation of income in a complete market. This section is influenced by the discussion in Cochrane (1991b).

First, note that complete markets must be Parato optimal²². This is well established in finance textbooks (see, for example, Huang and Litzenberger again) and so the proof is not given here. Intuitively it is reasonable. If one investor wants to instigate a series of trades amongst counterparties who are (at worst) indifferent against whether the trades are made or not and markets permit the trades then the trades will take place. The ultimate allocation will

²²“An allocation of state contingent claims is said to be *Parato optimal* or *Parato efficient* if it is feasible and if there do not exist other allocations which are feasible and can strictly increase at least one individual’s utility without decreasing the utility of others”, Huang and Litzenberger (1988) p.121, their italics.

thus end up being Parato optimal. Notice, though that incomplete markets need not provide Parato optimal allocations of claims as there may be trades that would increase the utility of some investors and decrease the utility of no investor, which might not be permitted through the incompleteness of the market.

The *social welfare function* is now applied to the complete market case. Suppose the economy has K agents each with utility U_k $k \in (1, K)$. Then the allocation of claims in this economy will be identical to one where the allocation of claims is made by maximising the social welfare function $\Upsilon(U_1, \dots, U_K)$ where:

$$\Upsilon(U_1, \dots, U_K) := \sum_{k=1}^K \lambda_k U_k$$

where $\lambda_k \geq 0 \forall k$. (see, for example, proposition 16.E.2 in Mas-Colell, Whinston and Green (1995)). Therefore, in a complete market, the action of the K individual agents can be modeled by maximising the linear social welfare function. To simplify the algebra assume that the economy is one period (see Cochrane (1991b) for the multiperiod proof). At time $t = 1$, there are S potential states with probability associated with each P_s $s \in (1, S)$ — it is assumed that there are homogeneous expectations. In a complete market, the individual optimisation problems of the individuals can be replaced by maximising the social welfare function subject to the budget constraints that $\sum_k c_{k0} = \sum_k y_{k0}$ and, for all s , $\sum_k c_{ks1} = \sum_k y_{ks1}$, where c_{k0}, c_{ks1} refer to the consumption at times 0 and time 1, should state s occur, of consumer

k . y_{k0}, y_{ks1} has analogous interpretation for endowment (income). There are, thus, $S+1$ potential budget constraints. Assuming that the utility U_k of each investor is time-separable and state independent, the maximisation problem becomes:

$$\max_{\{c_{k0}, c_{ks1}\}} \sum_{k=1}^K \sum_{s=1}^S \lambda_k [U_k(c_{k0}) + P_s U_k(c_{ks1}, 1)] - \mu_0 \sum_{k=1}^K [c_{k0} - y_{k0}] - \sum_{s=1}^S \mu_s \sum_{k=1}^K [c_{ks1} - y_{ks1}]$$

where the μ s represent the Lagrange multipliers. Maximising with respect to c_{k0}, c_{ks1} :

$$\begin{aligned} \lambda_k U'_k(c_{k0}) &= \mu_0 \quad \forall k \\ \lambda_k U'_k(c_{ks1}, 1) &= \mu_s / P_s \quad \forall k, s \\ \Rightarrow \frac{U'_k(c_{ks1}, 1)}{U'_k(c_{k0})} &= \frac{\mu_s}{P_s \mu_0} \quad \forall k, s \end{aligned} \tag{9}$$

Notice that the righthand side of this last equation is independent of k . Therefore the ratio of marginal utility of next period's consumption to the marginal utility of this period's consumption is fixed across investors which ever state occurs at $t = 1$.

Empirical tests of this full insurance hypothesis — that changes in consumption should be independent of the allocation of idiosyncratic endowment shocks — have been conducted by Mace (1991) (with a comment by Nelson (1994)), Cochrane (1991b) and Attanasio and Davis (1996). It is found that the allocation of income shocks does influence relative changes in household consumption. “In our view, the magnitude of the covariance between relative wages and consumption constitutes a spectacular failure of the hypothesis of between-group consumption insurance” (Attanasio and Davis

(1996), p.1259). Cochrane (1991b) finds that, in particular, involuntary job loss and long periods of sickness result in significantly lower growth rates of consumption than the aggregate (although, surprisingly, the duration of unemployment is not shown to be significant). This provides support for the informal linking of endowment shocks with unemployment that is made throughout this dissertation and the idea that complete market models may not accurately reflect the savings and consumption decisions of individuals.

5.2 The representative agent

Most tests of the Euler equation and related models assume that there is a representative agent whose preferences are well represented by power utility function $U(c_t) = \beta^t(c^{1-\gamma} - 1)/(1 - \gamma)$. This section provides a sufficiency condition for such a representative agent to exist²³. This result is given in sections 5.24–5.25 of Huang and Litzenberger (1988). The conditions that we will require are that the marginal utility for all agents $U'_k(c_{kt}) = \beta^t c_{kt}^{-\gamma}$ where β, γ are independent of k . In this case, substituting back into the last line of 9 gives:

$$\beta \left(\frac{c_{ks1}}{c_{k0}} \right)^{-\gamma} = A_s \quad \forall k, s$$

where $A_s := \mu_s/(P_s \mu_0)$ — a constant²⁴. Therefore, $c_{ks1} = (A_s/\beta)^{-(1/\gamma)} c_{k0}$

²³General papers on the existence of a representative agent are given by Rubinstein (1974) and, of particular relevance in this context, Constantinides (1982). Scheinkman (1989) provides a very clear example to demonstrate that assuming the existence of a representative agent is generally invalid in incomplete markets.

²⁴ μ_s/μ_0 has interpretation. From table 1 in chapter 2, it is known that the pricing kernel π_s can be represented by ϖ_s/P_s , where ϖ_s is the price of the pure security for state s . It will be shown below that, for the Euler equation, the pricing kernel is $U'(c_1, 1)/U'(c_0, 0)$.

or:

$$\begin{aligned} \sum_{k=1}^K c_{ks1} &= (A_s/\beta)^{-(1/\gamma)} \sum_{k=1}^K c_{k0} \quad \forall s \\ \Rightarrow \beta \left(\sum_{k=1}^K c_{ks1} / \sum_{k=1}^K c_{k0} \right)^{-\gamma} &= A_s \quad \forall s \end{aligned}$$

That is, the allocation of consumption across states and time would have been identical if the K individual agents had been modeled by a single representative agent whose marginal utility is the same as that of each of the individual agents. The sufficiency condition for the representative agent has thus been established.

5.3 The Euler equation for the representative agent

Assume the existence of the representative agent in an economy where there is only one consumption good. Utility of consumption is time-separable and state independent ²⁵; $u(c_1, c_2, \dots) = U(c_1, 1) + U(c_2, 2) + \dots$. There are n production units all producing the consumption good. Denote the output of the i th production unit at time t as d_{it} . The shares in these production units are traded ex-dividend. It is not possible to store this good (this is a “pure exchange economy”). Denote the proportion of the total holdings in the i^{th} production processes held by the representative agent at $t = 0$ and before any trading by z_i . As the agent must hold all assets $z_i = 1 \forall i$. At $t = 0$, after the production units have paid their dividends, the agent is, in

Therefore, $\mu_s/\mu_0 = \varpi_s$; the price of the pure security. See, again, section 5 of Huang and Litzenberger, who set $\mu_0 = 1$ and associate μ_s with the price of the pure security.

²⁵The extension to time non-separable utility is straight forward.

theory, able to trade these assets in order to maximise expected utility of consumption. Denote by Z_i the proportion of the total holdings in the i^{th} production processes held by the representative agent at $t = 0$ after trading. However, as she is the only investor in the world, $Z_i = z_i = 1 \forall i$.

Let p_i denote the price of asset i at time 0. Given that there is no storage potential, if we denote z, Z to be the $(1 \times n)$ vectors of holdings at time 0 before and after trading respectively, p to be the $(n \times 1)$ vector of prices and d_t to be the $(n \times 1)$ output vector, it is clear that the consumption of our investor at time 0 is $(z - Z).p + z.d_0$ and at time $t > 0$ is $Z.d_t$. So, the problem for the investor becomes:

$$\max_Z \quad U[(z - Z).p + z.d_0] + \sum_{t=1}^{\infty} \beta^t E_0(U[Z.d_t]) \quad \text{s.t. } z = Z = \mathbf{1}$$

where $\mathbf{1}$ is just a $(1 \times n)$ vector of 1's. By undertaking the constrained maximisation with respect to the n variables in Z and setting to zero, it is clear that we come up with n equations of the form:

$$-p_i U'[(z - Z).p + z.d_0] + \sum_{t=1}^{\infty} \beta^t E_0(d_{it} U'[Z.d_t]) = 0 \quad \forall i$$

Using the constraint that $Z = z = \mathbf{1}$:

$$p_i = \frac{\sum_{t=1}^{\infty} E_0(d_{it} U'[\mathbf{1}.d_t, t])}{U'[\mathbf{1}.d_0, 0]}$$

$\mathbf{1}.d_0$ is just $\sum_{i=1}^n d_{i0}$, or the total production of the physical good at time 0. Given that there is no storage in the economy, total production must

equal total consumption, so that $1 \cdot d_0 = c_0$, where c_0 is total consumption at time 0. Similarly, $1 \cdot d_t = c_t$, where c_t is total consumption at time t . We have now arrived at Lucas' key 1978 result in a multiperiod pure exchange economy with a representative agent where that agent has time separable, state independent utility:

$$p_i = \frac{\sum_{t=1}^{\infty} E_0(d_{it} U'[c_t, t])}{U'[c_0, 0]} \quad (10)$$

5.4 The Euler equation in incomplete markets

The proof of the Euler equation has so far only been demonstrated in the complete market case. As this thesis concentrates on incomplete markets, it is important to establish the general version of this equation. Consider the K investors in the economy who have time separable and state independent utility of consumption $u_k(c_0^k, c_1^k, \dots) = U_k(c_0^k, 0) + U_k(c_1^k, 1) + \dots$. In addition to the n production units, each agent receives an endowment y_{kt} at time t that cannot be traded. If $y_{kt} = 0 \forall k, t$, then we return to the complete market case. Suppose that at time $t = -1$ there is a feasible distribution of shares in each production process i amongst the k investors. When $t \in [-1, 0)$, trading takes place in financial markets to establish a price p_i for each asset i . At time $t = 0$, investor k has holdings z_{ki} in asset i under the constraint that $\sum_k z_{ki} = 1 \forall i$. If the market is in equilibrium at time $t = 0$, then each investor is "happy" with her holdings z_{ki} and has no incentive to change to any other holding Z_{ki} where $Z_{ki} \neq z_{ki}$. That is (in vector notation) $\mathbf{Z}_k = \mathbf{z}_k$

must be a solution to:

$$\max_{\mathbf{Z}_k} U_k[(\mathbf{z}_k - \mathbf{Z}_k)\mathbf{p} + \mathbf{z}_k\mathbf{d}_0 + y_{k0}] + \sum_{t=1}^{\infty} E_0[U_k[\mathbf{Z}_k\mathbf{d}_t + y_{kt}]] \quad \text{s.t. } \mathbf{z}_k = \mathbf{Z}_k$$

Undertaking the constrained optimisation:

$$-p_i U'_k[(\mathbf{z}_k - \mathbf{Z}_k) \cdot \mathbf{p} + \mathbf{z}_k \cdot \mathbf{d}_0 + y_{k0}] + \sum_{t=1}^{\infty} \beta^t E_0(d_{it} U'_k[\mathbf{Z}_k \cdot \mathbf{d}_t + y_{kt}]) = 0 \quad \forall i$$

As $\mathbf{Z}_k = \mathbf{z}_k$ must be a solution to this optimisation:

$$p_i = \frac{\sum_{t=1}^{\infty} E_0(d_{it} U'_k[\mathbf{z}_k \cdot \mathbf{d}_t + y_{kt}, t])}{U'_k[\mathbf{z}_k \cdot \mathbf{d}_0 + y_{k0}, 0]}$$

Now $\mathbf{z}_k \cdot \mathbf{d}_0 + y_{k0} = c_0^k$ and $\mathbf{z}_k \cdot \mathbf{d}_t + y_{kt} = c_t^k$, so the general Euler equation in an incomplete market is:

$$p_i = \frac{\sum_{t=1}^{\infty} E_0(d_{it} U'_k[c_t^k, t])}{U'_k[c_0^k, 0]} \quad (11)$$

This is the (stochastic) Euler equation and is, in many ways, the central theoretical model of this thesis. As outlined by Grossman and Shiller (1981) (their italics, p. 223):

The theory of asset returns embodied in ... (the Euler equation) ... is very powerful because it can be applied so generally. It holds for *any* asset, or portfolio of assets. It holds for *any* individual consumer who has the option of investing in stocks (even if he chooses not to hold stocks) and thus it must hold for aggregate consumption so long as some peoples' consumption is well represented by the aggregate consumption. It holds even if the individual's choices regarding other assets are constrained ... The model holds for any time period ...

From above, we know that all equilibrium asset pricing models can be interpreted within the context of the fundamental theorem of asset pricing. That is, there is some strictly positive π such that, in a one period setting, for all i , $p_i = E[d_i\pi]$. From the Euler equation we know that $p_i = E[d_i U'_k(c_1^k, 1)/U'_k(c_0^k, 0)]$. The Euler equation is, therefore, easily reconciled with the fundamental theorem: $\pi \equiv U'_k(c_1^k, 1)/U'_k(c_0^k, 0)$. So, the association of the pricing kernel with the ratio of marginal utilities and the fact that this kernel will only be uniquely defined if the market is complete has now been established.

5.5 Towards the CCAPM

So far the Euler equation has been established both in complete and incomplete markets and placed in the context of the fundamental theorem of asset pricing. So, equations 2–6 hold in an Euler equation world with π_t replaced by $U'_k(c_t^k, t)/U'_k(c_{t-1}^k, t-1)$. In the complete market case, the k s can be “dropped”, while in the incomplete market case they cannot be excluded. This subsection aims to introduce the consumption CAPM and link it in to the Euler equation. By taking a discrete time approach, the key assumptions that underlies that CCAPM are readily apparent.

Remembering that $\pi = U'_k(c_1^k, 1)/U'_k(c_0^k, 0)$ for all k , equation 3 becomes $0 = E[(r_i - r_j)U'(c_1^k, 1)]$, which again holds for all k . Let $\bar{c}^k := E[c_1^k]$ and $\tilde{c}^k := c_1^k - \bar{c}^k$. So, $0 = E[(r_i - r_j)U'_k(\bar{c}^k + \tilde{c}^k, 1)]$ for all k . Taking a Taylor’s series expansion:

$$\begin{aligned}
E[r_i - r_j] &= -\frac{U_k''(\bar{c}^k, 1)}{U_k'(\bar{c}^k, 1)}E[\tilde{\epsilon}^k(r_i - r_j)] - \frac{U_k'''(\bar{c}^k, 1)}{2U_k'(\bar{c}^k, 1)}E[(\tilde{\epsilon}^k)^2(r_i - r_j)] - \dots \\
\Rightarrow E[r_i] - r_f &= \frac{-U_k''(\bar{c}^k, 1)\bar{c}^k}{U_k'(\bar{c}^k, 1)}E\left[\frac{\tilde{\epsilon}^k}{\bar{c}^k}r_i\right] + o(\tilde{\epsilon}^k)
\end{aligned}
\tag{12}$$

In order to derive the CCAPM, this equation needs to be reduced in two ways. First, the $o(\tilde{\epsilon}^k)$ term needs to be negligible. Second, individual consumption/utilities must be replacable with the consumption and utility of a representative agent. Take these in turn. The $o(\tilde{\epsilon}^k)$ term will be zero if utility is quadratic. It will be negligible if the consumption of all consumers is smooth. With regard to replacing individual consumption with aggregate consumption, this can certainly occur if the market is complete. It can also occur if the market is incomplete in the quadratic utility case provided that $\bar{c}^k, U_k(\cdot)$ are independent of k — in other words, if investors are “*ex-ante* homogeneous”. $\tilde{\epsilon}^k$ may still be k -dependent so there is no representative agent. In this case we can sum the left and right hands sides over k and, because covariances are additive, the individual $\tilde{\epsilon}^k$ terms can be replaced with aggregate consumption uncertainty. That the CCAPM will hold under assumptions of quadratic utility and incomplete markets with *ex-ante* homogeneous consumers was shown initially by Mankiw (1986)²⁶. Alternatively the assumption that all investors have smooth consumption and are *ex-ante* homogeneous will also give the CCAPM. A more formal derivation

²⁶In the next two chapters the role that the third and fourth derivative of the utility function plays in determining equilibrium asset prices in incomplete markets will be discussed in detail. This is the first indication that differences between complete and incomplete market models with *ex-ante* homogeneous investors will only occur with non-quadratic utility functions.

of this result will be given later in the chapter. Notice that in the type of model that will be developed in this thesis — incomplete markets, power utility and endowment shocks — the CCAPM will not follow because the $o(\tilde{\epsilon}^k)$ terms will be k -dependent and will not be negligible. Therefore, even though investors will be ex-ante homogeneous, it is not possible to sum over k and replace individual consumption with aggregate consumption. So, even if aggregate consumption is smooth, provided that individual consumption is not and provided that markets are incomplete, the CCAPM will not hold.

5.6 The CCAPM in discrete time

Having loosely linked the CCAPM to the Euler equation, this subsection provides a number of more formal discrete-time developments of the CCAPM. The theoretical robustness of the model to changes in underlying assumption that were discussed above are then demonstrated. While similar issues are discussed in continuous time, the papers in this area are less intuitive. So understanding the role of “smoothness” of consumption in the discrete time development of the CCAPM should help provide understanding for later sections. Throughout this section there is a representative agent: that is, this is a complete markets development of the CCAPM.

5.6.1 The CCAPM under bivariate normality

An early proof of the CCAPM was provided by Rubinstein (1976) under the assumptions of a representative agent and bivariate normality of consumption and asset returns in discrete time. Breeden and Litzenberger (1978)

extended this result to heterogeneous investors in a complete market — see the discussion on aggregation given above. The key to the result is that, if variables \tilde{x}, \tilde{y} are bivariate normal and $f(\cdot)$ is a differentiable function then $\text{Cov}[\tilde{x}, f(\tilde{y})] = E[f'(\tilde{y})]\text{Cov}[\tilde{x}, \tilde{y}]$. So, if r_i and c_1 (consumption here is for the representative agent and so the k superscript can be dropped) are bivariate normal for all i and letting r_c denote the return to a portfolio of assets (perhaps, but not necessarily, the portfolio of assets most correlated with aggregate consumption) then, rearranging equation 5 gives:

$$E[r_i] - r_f = E[r_c - r_f] \frac{\text{Cov}(r_i, c_1)}{\text{Cov}(r_c, c_1)} \quad (13)$$

This is a version of the consumption CAPM.

5.6.2 The CCAPM under quadratic utility

That the CCAPM follows from the assumption of quadratic utility $U(c_1, 1) = \beta(ac_1 - bc_1^2)$ in a complete market is also easily shown. In fact, by substituting this utility function into equation 5, the CCAPM, equation 13, follows direct. That the CCAPM follows under quadratic utility is particularly important within continuous time for Grossman and Shiller (1982) who exploit the “local linearity” of marginal utility for differentiable utility functions to prove the CCAPM under fairly general conditions.

5.6.3 The CCAPM in discrete time under isoelastic utility, log-normal consumption & normal asset returns

Define $C_1 := \ln(c_1)$ and assume that C_1 is normal — that is, aggregate consumption is lognormal. If the representative agent exists and has power utility so that $U'(c_1, 1) = \beta c_1^{-\gamma}$, then equation 5 can be rearranged to give:

$$\begin{aligned} E[r_i] - r_f &= E[r_c - r_f] \frac{\text{Cov}[r_i, (\exp[C_1])^{-\gamma}]}{\text{Cov}[r_c, (\exp[C_1])^{-\gamma}]} \\ &= E[r_c - r_f] \frac{\text{Cov}(r_i, \exp[-\gamma C_1])}{\text{Cov}(r_c, \exp[-\gamma C_1])} \end{aligned}$$

Assume that r_i is normal and let $f(\tilde{y}) = \exp[-\gamma \tilde{y}]$. Remembering that C_1 is normal by assumption and applying the result that, for normal \tilde{x}, \tilde{y} , $\text{Cov}[\tilde{x}, f(\tilde{y})] = E[f'(\tilde{y})]\text{Cov}[\tilde{x}, \tilde{y}]$:

$$\begin{aligned} E[r_i] - r_f &= E[r_c - r_f] \frac{\text{Cov}(r_i, C_1)}{\text{Cov}(r_c, C_1)} \\ &= E[r_c - r_f] \frac{\text{Cov}(r_i, \Delta \ln c)}{\text{Cov}(r_c, \Delta \ln c)} \end{aligned}$$

which is an alternate (and more usual) version of the CCAPM where covariances are taken with respect to log consumption. Here $\Delta \ln c := \ln(c_1) - \ln(c_0)$ and the last line follows as $\text{Cov}[r_i, \ln(c_0)] = 0$ as c_0 is known with certainty.

5.6.4 The CCAPM: an approximation under isoelastic utility and smooth consumption

It has just been shown that if aggregate consumption is lognormal in a complete market and the returns on assets are normal then the CCAPM follows direct. It can be shown that if these assumptions are weakened so that

consumption is smooth and removing all restrictions on the distribution of asset returns, then the CCAPM follows as an approximation under isoelastic utility. If utility has power form, then:

$$\begin{aligned}\frac{U'(c_1, 1)}{U'(c_0, 0)} &= \beta \left(\frac{c_1}{c_0}\right)^{-\gamma} \\ &= \beta (\exp \Delta \ln c)^{-\gamma} \\ &\approx \beta (1 - \gamma \Delta \ln c)\end{aligned}\tag{14}$$

Terms in $o(\Delta \ln c)$ have been ignored in the approximation, which follows as $\exp(\delta) = 1 + \delta + o(\delta)$ and, from the binomial theorem, $(1 + \delta)^x = 1 + x\delta + o(\delta)$. Substituting the ratio of marginal utilities given in equation 14 into equation 5:

$$E[r_i] - r_f = E[r_c - r_f] \frac{\text{Cov}[r_i, \Delta \ln c]}{\text{Cov}[r_c, \Delta \ln c]}$$

6 The CCAPM in continuous time

This section gives a formal, continuous-time derivation of the CCAPM as derived originally by Breeden (1979). This section sticks closely to the Breeden proof and, indeed, the notation in this section differs somewhat from that used elsewhere in the thesis in order to to be comparable with the original paper.

The proof is based on a dynamic programming approach as initially advanced by Merton (1969), Merton (1971)²⁷. By looking at a partial equilib-

²⁷See Hakansson (1970) for another important early paper analysing consump-

rium economy²⁸ with asset returns and consumption assumed to follow Itô processes driven by m state variables, the Intertemporal CAPM (ICAPM) of Merton (1973) is proven. By concentrating on the fact that utility is state-independent, it is shown that the multi-beta ICAPM can be reduced to the single beta CCAPM with no further assumptions. This is the insight of Breeden.

Assumption 1 *There are K investors (not necessarily homogeneous) within the economy. At time t , each investor, k , is given an initial allocation of wealth W^k . The investor receives no further endowment of the consumption good.*

Assumption 2 *There is only one physical good in the economy. Utility of consumption can therefore be expressed in terms of this single physical good²⁹.*

Assumption 3 *Utility is time-additive and of fixed form. So, the utility of consumption before the economy terminates at known time T is given by $\int_t^T U(c(r), r)dr$. It is assumed throughout that the utility function is state independent³⁰.*

tion/savings decisions. While the dynamic programming approach has now been largely superseded by the martingale technique of Cox and Huang (1989), it is still a satisfactory method for analysing many types of continuous-time savings and consumption decisions. See, for example, Karatzas, Lehoczky, Sethi and Shreve (1986) who use this approach to solve a very general class of investment problems and Svensson and Werner (1993) who consider the investment problem under exponential utility in the presence of non-tradable risk.

²⁸The partial equilibrium assumption imposes structure on asset price and consumption movements. This partial equilibrium approach is in contrast to the general equilibrium models of, for example, Cox, Ingersoll and Ross (1985).

²⁹Breeden (1979) does develop the CCAPM in a multi-good economy but this is outside the scope of this analysis.

³⁰This is required for the consumption CAPM. The intertemporal version can be derived

Assumption 4 *There exists a riskless asset paying an exogenous rate of return r at each instant³¹.*

Assumption 5 *There are n traded risky assets in the economy that are influenced by m state variables. The evolution of these state variables, s_i , and the asset prices, P_i can be described by exogenous diffusion processes:*

$$\begin{aligned} ds_i &= f_i dt + g_i dq_i \\ dP_i/P_i &= \alpha_i dt + \sigma_i dz_i \end{aligned}$$

where q_i, z_i are Weiner processes. Denote the instantaneous correlation between dq_i and dq_j by v_{ij} .

Assumption 6 *Asset price evolve over time in a way prescribed by the m state variables. Denote the covariance between z_i, z_j by σ_{ij} and let the correlation coefficient between z_i, q_j be denoted by η_{ij} .*

The problem that the investor faces is as follows. She must make her wealth W^k last for the remainder of her life and may derive a “utility of bequest”, $B^k[W(T), T]$ if leaving an amount $W(T)$ to her benefactors when she dies at a known (non-stochastic) future time T . She is therefore trying to maximise two things: how much to consume c^k at each time and what percentage of her remaining wealth w^k (a $(n \times 1)$ vector) should she place in each of the n risky assets (with the remainder going to the riskless asset).

under the more general condition of state dependent utility. The reader is referred to the reprint of Merton (1973) in Merton (1992) (Chapter 15) for the relevant extension.

³¹Again, Breeden derives the CCAPM in an environment without a riskless asset and the interested reader is referred to the original paper.

At any instant in time, her budget constraint (dropping the k superscripts) is that:

$$\begin{aligned}
 W(t + \delta t) &= W(t) - c\delta t + (1 - \mathbf{w}^T \cdot \mathbf{1})Wr\delta t + \sum_{i=1}^n w_i W dP_i \\
 \Rightarrow dW &\approx -cdt + (1 - \mathbf{w}^T \cdot \mathbf{1})Wr dt + \sum_{i=1}^n w_i W (\alpha_i dt + \sigma_i dz_i) \\
 &= \left[\sum_{i=1}^n w_i (\alpha_i - r) + r \right] W dt + \sum_{i=1}^n w_i W \sigma_i dz_i - cdt
 \end{aligned} \tag{15}$$

where c is the “rate of flow of consumption” at that instant (so that total consumption in time period $(t, t + \delta t) \approx c\delta t$) and $\mathbf{1}$ is the usual $(n \times 1)$ vector of 1s. We can formulate the investor’s problem as

$$\max_{(c, \mathbf{w})} E_t \left[\int_t^T U(c, r) dr + B[W(T), T] \right] =: J(W^k, \mathbf{s}, t) \tag{16}$$

Here $J(W^k, \mathbf{s}, t)$ denotes the total utility (including bequest) made from making the optimal choices \mathbf{w}, c . Equation 16 can be solved using a dynamic programming approach: see appendix 9.5. The optimal holding of the n risky assets at time t is given by (re-introducing the k superscript)³²:

$$W^k \mathbf{w}^k = \frac{-J_{WW}^k}{J_{WW}^k} \mathbf{V}_{aa}^{-1} (\mathbf{a} - r\mathbf{1}) - \mathbf{V}_{aa}^{-1} \mathbf{V}_{as} \frac{\mathbf{J}_s W^k}{J_{WW}^k} \tag{17}$$

Here \mathbf{a} is a $(n \times 1)$ vector whose i^{th} element is α_i , \mathbf{V}_{aa} represents the $(n \times n)$ returns covariance matrix for the n different assets and \mathbf{V}_{as} represents

³²Svensson and Werner (1993) show that, if there is one investor (so the k superscripts are redundant) who receives untradable income $\mathbf{y}(t)$ which describes an Itô process, then we need to add an additional term $-\mathbf{V}_{aa} \mathbf{V}_{a(y)}$ to the right hand side of this equation, where $\mathbf{V}_{a(y)}$ is a $(n \times 1)$ vector of covariances between asset returns and the nontradable income.

an $(n \times m)$ matrix of covariances between asset returns and movements in the state variables. Otherwise the general notation x_y denotes the partial derivative of x with respect to y . So, for example, \mathbf{J}_{sW}^k is a $(m \times 1)$ vector whose i^{th} element is $J_{s_i W}^k$, $\partial J^k / \partial s_i \partial W^k$. From the envelope condition, $J_W^k = U_c^k$, $J_{WW}^k = c_W^k U_{cc}^k$ and $\mathbf{J}_{sW}^k = \mathbf{c}_s^k U_{cc}^k$ ³³, where \mathbf{c}_s^k is a $(1 \times m)$ vector whose i^{th} element is $c_{s_i}^k$. Let T^k represent the absolute risk tolerance for investor k : $(-U_c^k / U_{cc}^k)$. Equation 17 becomes:

$$\begin{aligned} W^k \mathbf{w}^k &= T^k \frac{\mathbf{V}_{aa}^{-1}}{c_W^k} (\mathbf{a} - r\mathbf{1}) - \mathbf{V}_{aa}^{-1} \mathbf{V}_{as} \frac{\mathbf{c}_s^k}{c_W^k} \\ \Rightarrow T^k (\mathbf{a} - r\mathbf{1}) &= \mathbf{V}_{aa} \mathbf{w}^k W^k c_W^k + \mathbf{V}_{as} \mathbf{c}_s^k \end{aligned} \quad (18)$$

This equation will form the basis for both the intertemporal CAPM of Merton (1973) and the consumption CAPM of Breeden (1979).

6.1 The Intertemporal Capital Asset Pricing Model

Merton's original 1973 paper proves the ICAPM with one state variable (the riskless rate) and obtains a specific case of the ICAPM. This is also the main textbook treatment (see, for example, p.280 et. seq. of Ingersoll (1987)).

The reprint of Merton (1973) in Merton (1992) produces the general pricing relationship in a world with m state variables and with state dependent utility

³³If $U(\cdot) = U(c, s, t)$ then $J_{s_i W}^k = c_{s_i}^k U_{cc}^k + U_{cs_i}$ and the proof of the CCAPM no longer follows. To quote Stephen Ross (1989) "Breeden's analysis of the Merton model exploited the observation that if the local utility function was not dependent on the state of nature other than through the dependence of the optimal consumption choice on the state, then along an optimal path the marginal utility of wealth would depend only on consumption. The key to the result is the state independence of the utility function" (p.88). The reprint of Merton (1973) in Merton (1992), however, shows that the ICAPM will still hold true.

of consumption $U(\cdot)$. Here, a “middle” route is followed where there are many states (initially) but the utility of consumption function is state independent. Reduction to one state variable occurs at the last stage to get tractability without over-elaborating on the algebra. Define $\mathbf{H}_s := \sum_k \mathbf{c}_s^k / c_W^k$. Similarly define $T := \sum_k T^k / c_W^k$ and $\mathbf{w}W := \sum_k \mathbf{w}^k W^k$, the total money invested in each asset. Then, summing equation 18 over k and rearranging:

$$\mathbf{w}W = T\mathbf{V}_{aa}^{-1}(\mathbf{a} - r\mathbf{1}) - \mathbf{V}_{aa}^{-1}\mathbf{V}_{as}\mathbf{H}_s$$

Assume that assets $n - m + 1$ to n are perfectly correlated with the m state variables. That is, asset $n - m + i$ is perfectly correlated with state variable $i \in [1, m]$. Then, for the return of any asset r_j , $\text{Cov}(r_j, s_i) = \text{Cov}(r_j, r_{n-m+i})\text{Std}(s_i)/\text{Std}(r_{n-m+i}) = \text{Cov}(r_j, r_{n-m+i})g_i/\sigma_{(n-m+i)}$. So, \mathbf{V}_{as} is the last m columns of \mathbf{V}_{aa} multiplied by an $(m \times 1)$ vector whose i th element is given by $g_i/\sigma_{(n-m+i)}$. Therefore, $\mathbf{V}_{aa}^{-1}\mathbf{V}_{as}$ is the last m columns of a diagonal matrix whose non-zero element in the i th column (of the truncated matrix) is $g_i/\sigma_{(n-m+i)}$. This $(n \times m)$ matrix will be denoted by \mathbf{K}_{as} . Rearranging the previous offset equation:

$$\mathbf{a} - r\mathbf{1} = \frac{\mathbf{V}_{aa}\mathbf{L}_{a(s)} + \mathbf{V}_{M(a)}W}{T} \quad (19)$$

$\mathbf{L}_{a(s)} := \mathbf{K}_{as}\mathbf{H}_s$ is a $(n \times 1)$ vector whose elements are zero up to, but not including, the $(n - m + 1)$ th and non-zero thereafter. $\mathbf{V}_{M(a)} = \mathbf{V}_{aa}\mathbf{w}$ is a $(n \times 1)$ vector of covariances between the market portfolio and the individual assets. Left multiply the this equation by \mathbf{w}^T , the weightings of each

asset in the market portfolio. Then the left hand side becomes the excess expected return to the market portfolio r_M (even if there is non-zero aggregate investment in the riskfree asset) and:

$$E[r_M - r_f] = \frac{\mathbf{V}_{\mathbf{M}(\mathbf{a})}^{\mathbf{T}} \mathbf{L}_{\mathbf{a}(\mathbf{s})} + \text{Var}(r_M)W}{T}$$

In order to get an equilibrium asset pricing relationship there are $m + 1$ elements to identify: then m non-zero elements in $\mathbf{L}_{\mathbf{a}(\mathbf{s})}/T$ and W/T . The $m + 1$ variables that will be substituted in for these element are the expected returns on the m assets that are perfectly correlated with the state variables and the return to the market. We are thus in a position to create an equilibrium asset pricing relationship. Substituting in for W/T in equation 19 and coming out of matrix notation we get:

$$E[r_i - r_f] = \frac{\sigma_{Mi}}{\sigma_M^2} E[r_M - r_f] + \sum_{j=1}^n \left[\sigma_{ij} - \frac{\sigma_{Mi}\sigma_{Mj}}{\sigma_M^2} \right] \frac{L_{j(\mathbf{s})}}{T}$$

where $L_{j(\mathbf{s})}$ is the j th element of $\mathbf{L}_{\mathbf{a}(\mathbf{s})}$ and sigmas represent variances and covariances of asset returns. In general this is complex to solve analytically but will lead to the general ICAPM pricing relationship. In particular, if $m = 1$, it is simple (but long-winded) to show that this equation can be rearranged to give equation 34 of Merton (1973) — the main result of that paper. It is also clear from this equation that if $m = 0$ then the CAPM follows. Thus it is possible to construct the CAPM in a continuous-time economy if there is a constant opportunity set³⁴.

³⁴For an excellent discussion of the links between dynamic and static equilibrium models

6.2 The Consumption Capital Asset Pricing Model

6.2.1 The Breeden proof

Return to equation 18. Consider the term $V_{aa}w^k W^k$. The element V_{aa} is the covariance matrix of asset returns. The term $w^k W^k$ tells us how investor k has distributed her wealth across assets. It is therefore clear that $V_{aa}w^k W^k = V_{aWk}$. That is it represents a $(n \times 1)$ vector whose i^{th} element is the covariance between the returns of asset i and the wealth of investor k . Rewriting equation 18:

$$T^k(a - r1) = c_W^k V_{aWk} + V_{as}c_s^k \quad (20)$$

Next, look at V_{ack} , that is the $(1 \times n)$ vector of covariances between the returns of the assets and the change in consumption of investor k . Take the i^{th} element of this: $Cov(r_i, c^k)$. We know that $c^k = c(W^k, s)$, so, using the chain rule $dc^k = c_W^k dW^k + \sum_{i=1}^m c_{s_i}^k ds_i$. So, the i^{th} element of $V_{ack} = Cov(r_i, c_W^k W^k + \sum_{i=1}^m c_{s_i}^k s_i)$. Separating out the covariances and returning to matrix notation, it is clear that:

$$V_{ack} = c_W^k V_{aWk} + V_{as}c_s^k \quad (21)$$

Notice that the right hand sides of equations 20 and 21 are the same.

Therefore, equating the left hand sides:

the reader is referred to Ross (1989).

$$\mathbf{V}_{\mathbf{a}c\mathbf{k}} = T^k(\mathbf{a} - r\mathbf{1}) \quad (22)$$

The next step is to aggregate this across the economy. That is, sum the left and right hand sides of equation 22 across all k . Because of the additive nature of covariances, $\sum_k \mathbf{V}_{\mathbf{a}c\mathbf{k}} = \mathbf{V}_{\mathbf{a}c}$ where c is total consumption in the economy. Letting $T := \sum_k T^k$:

$$\mathbf{a} - r\mathbf{1} = T^{-1}\mathbf{V}_{\mathbf{a}c} \quad (23)$$

Multiply and divide the right hand side of equation 23 by C . The i^{th} element of $\frac{c}{c}\mathbf{V}_{\mathbf{a}c} = c/c\text{Cov}(r_i, dc) = c\text{Cov}(r_i, dc/c) \approx c\text{Cov}(r_i, \Delta \ln c)$. We can therefore clearly rewrite equation 23 as:

$$\mathbf{a} - r\mathbf{1} = \frac{c}{T}\mathbf{V}_{\mathbf{a}\Delta \ln c} \quad (24)$$

This is an important representation of the CCAPM, a version of which was derived in discrete time in section 5.5 but perhaps which has not been sufficiently emphasised to date. All the discrete time versions of the CCAPM given above could be presented in a form similar to this. c/T is the coefficient of relative risk aversion for the representative agent. So, the ex-ante risk premium for any asset in equilibrium is just the coefficient of relative risk aversion of the representative agent multiplied by the covariance between the asset's returns and changes in log consumption. This will be important when discussing the equity premium puzzle in the next chapter. Returning to the previous offset equation, we can substitute in for c/T by looking at returns

on portfolio with returns r_c . Substituting back in we get the consumption CAPM in its “usual” form:

$$E[r_i - r] = E[r_c - r] \frac{\text{Cov}[r_i, \Delta \ln c]}{\text{Cov}[r_c, \Delta \ln c]}$$

6.2.2 Exploiting local linearity

While this analysis gives the original derivation of the CCAPM as given by Breeden (1979). Grossman and Shiller (1982), however, exploit Itô’s lemma to provide a simpler derivation of the result. The exposition here does not consider the aggregation of consumers but assumes that a representative agent exists. A more formal analysis of the aggregation process is given in the original paper. Let V_{it} represent the value of any asset i at time t . Assume that (V_{it}, c_t) is an Itô process, then, $Z_{it} := U'(c_t)V_{it}$ develops according to:

$$dZ_{it} = U'(c_t)dV_{it} + U''(c_t)[V_{it}dc_t + dc_t dV_{it}] + 0.5U'''(c_t)V_{it}(dc_t)^2$$

The Euler equation tells us that the expectation of the left hand side is the same for any two assets i, j . So, equating the expectation of the right hand side for i, j :

$$E \left[\frac{dV_{it}}{V_{it}} - \frac{dV_{jt}}{V_{jt}} \right] = \frac{-U''(c_t)c_t}{U'(c_t)} \text{Cov} \left[\frac{dc_t}{c_t}, \frac{dV_{it}}{V_{it}} - \frac{dV_{jt}}{V_{jt}} \right] \quad (25)$$

Letting asset j be the riskless asset, the CCAPM follows direct.

6.3 The CCAPM and ICAPM with nonmarketable assets

Formal conditions on the primitives of the economy which will guarantee the existence of an equilibrium supporting the CCAPM are provided by Duffie and Zame (1989). Within a continuous-time environment, the conditions are, essentially, that utilities are smooth, monotonic increasing and concave with marginal utility tending to infinity as consumption tends to zero from above, aggregate endowment is an Itô process and markets are complete.

Grossman and Shiller (1982) allow investors to hold assets whose returns might not describe Itô processes and which might be traded with transaction costs³⁵. They show that, provided the overall consumption pattern of each individual follows an Itô process, the CCAPM still holds (over short time intervals) for the subset of assets whose returns are Itô and which can be traded costlessly. David Brown (1988) examines the CCAPM and ICAPM under the assumption that a subset of the population receives continuously changing nonmarketable income. He also places liquidity constraints on these individuals so that they cannot have a portfolio of marketable assets with negative value — thus distinguishing his paper from that of Grossman and Shiller (1982). He shows that, under logarithmic utility, these constrained investors may find it optimal to have zero wealth (no investment in marketable assets) and to consume “hand-to mouth” their nonmarketable income. In this case, he argues, the CCAPM will no longer hold in its original form. Instead, a

³⁵They also allow for heterogeneous information, but this is of less interest to us here.

variation of the CCAPM is derived where consumption betas are calculated against the aggregate consumption of the subset of individuals with positive wealth. Essentially his proof is as follows. Consider an economy with no constrained individuals. Then the CCAPM holds in its original form. Now add a constrained investor to the economy whose optimal policy is to consume hand-to-mouth. This investor will not affect asset prices as she will have no wealth to invest in marketable assets and is constrained from selling short. However, the introduction of this agent changes aggregate consumption and so, possibly, the consumption betas of the individual assets. The CCAPM cannot hold both in the presence and absence of this individual and the pricing relationship is therefore refuted. Brown points out, however, that the ICAPM still holds. The reason for this is as follows. In the CCAPM, aggregation is done across consumption. Introducing constrained investors whose optimal policy is to hold no wealth affects this aggregation process. In the ICAPM, however, aggregation is done across wealth. Given that the constrained individuals who consume hand-to-mouth have no wealth, introducing such agents does not affect this aggregation process. The proof of the ICAPM proceeds unhindered. The proof of the CCAPM requires that aggregation is carried out across investors with positive wealth only.

A paper of great relevance for the discussions to come is Back (1991)³⁶

³⁶See also Aase (1993), who considers a similar problem in a representative agent environment. Aase (1996) extends the earlier paper by considering cases when the jump-diffusion processes of asset returns and aggregate consumption are assumed to be Gaussian inverse Gaussian and inverse Gaussian respectively. In this case, there is a closed form "Consumption CAPM" (his equations 3.38, 3.39), which is, of course, significantly different from the CCAPM given above for returns that are continuous.

who describes an economy where individual optimal consumption paths and asset returns are assumed to follow jump-diffusion processes.

The risk premium of a security can be divided into two parts: the premium for the continuous part of the return and the premium for the jump part. Each premium is proportional to its covariance with the corresponding part (continuous or jump) of the state price density process. The covariance between the jump parts of the security price and state price processes is zero unless jumps occur simultaneously. Thus, if the state price density does not jump simultaneously with the security, then the jump risk of the security is unpriced.

Back (1991), p.372

The amended version of the CCAPM in the case of jump-diffusion returns and consumption paths is given in his theorem 3. The risk premium on the continuous part of asset returns is determined by its covariance with the continuous part of *aggregate* consumption. To stress: even in an incomplete market we can use the continuous part of aggregate consumption to price the continuous part of asset returns. This result was hinted at much earlier in the chapter in section 5.5. “The assumption of complete markets is not necessary, however, to obtain a formula for the continuous risk premium in terms of aggregate consumption, because the linear relations for the various investors can be aggregated” (*ibid.* footnote 18). The risk premium on the discontinuous parts of an asset’s return is determined its covariance with a variable η . “...notice that the process, η is not defined simply in terms of jumps in the aggregate consumption rate, but rather depends in a complicated way on the jumps in the various investor’s consumption rates” (*ibid.* p.387). As in an incomplete market the optimal consumption path of each investor need not be perfectly correlated risk premia will depend on the

distribution of consumption jumps amongst investors. This is first evidence presented in this thesis to suggest that introducing low probability, high impact uninsurable shocks to individual endowments will make significant difference to equilibrium asset prices. Jarrow and Rosenfeld (1984) (US) and Bentzen and Sellin (1997) (other major markets) provide empirical evidence showing that there are jumps in the return to market indices, suggesting that this may provide a potential explanation to the equity premium puzzle³⁷.

An important implication of the Kerry Back's work is that, if asset prices and optimal consumption paths are jump-diffusion (with correlated jumps), estimated risk premia estimated from the CCAPM of Breeden (1979) will be "too low" (or, alternatively, estimates of risk aversion will be "too high"). Therefore, as observed by Aase (1993), Aase (1996), jump-diffusion asset returns and aggregate consumption may form a (partial) solution to the equity premium puzzle. Back's result takes this intuition one stage further by showing that jumps in individual (not aggregate) consumption paths and asset returns may form the solution to the puzzle. This issue will be addressed in much greater detail in later chapters.

Grossman and Laroque (1990) assume that utility comes from a durable good, K_t , which can be bought and sold and otherwise depreciates in a deterministic manner $dK_t = -\alpha K_t$. To change holdings in the durable good, an investor must sell her existing durable good with transaction costs and

³⁷Although both develop a theoretical framework that states that jumps in individual share price processes will only be priced if they are correlated with these market level jumps. This theoretical paradigm is less general than that of Back and Aase as there are no endowments, no state variables and markets are complete in their models.

repurchase a durable good with the required value. Clearly the introduction of transaction costs will mean that investors will want to change holdings in the durable good rarely so that K_t will describe a jump and depreciation (no diffusion) process. Changes in investor wealth otherwise take place through holdings in marketable assets (which are assumed to follow Itô processes) and the riskless asset. In this case, it can be shown that the CCAPM will not hold but that the CAPM will. Intuitively this is reasonable. The consumption good follows a jump and deterministic depreciation process, asset returns follow a diffusion process. As pointed out by Back (1991) (p.378), the covariance between a diffusion and a jump process is necessarily zero. Therefore all assets have zero consumption beta and the CCAPM cannot hold. The CAPM will, though, hold. Investors are trying to minimise the time τ , subject to their willingness to bear risk, before it is rational, under transaction costs, to sell their existing durable good and trade up. Therefore, over a short investment horizon $(t, t + 1)$, investors will, subject to risk aversion, be trying to maximise financial wealth at $t + 1$. That is, at time t they will be behaving in financial markets as if they were maximising $v(W_{t+1})$ for some utility function $v(\cdot)$ and level of financial wealth W . Using the “local linearity” condition of Grossman and Shiller (1982) and assuming that the interval $(t, t + 1)$ is sufficiently short, if asset returns are Itô then utility of financial wealth will behave as though it were quadratic whatever its “real” form. It is well known that quadratic utility implies the CAPM. Therefore, the CAPM follows as a consequence of investors behaving as though they

were maximising their utilities of financial wealth at the next instant and through the local linearity condition that follows from the diffusion of asset returns.

So, as can be seen, there are a number of theoretical developments to the CCAPM since Breeden (1979) that are relevant to this dissertation. In general, the CCAPM holds if either consumption patterns or asset returns are diffusion process but does not hold if there are jump elements. The exception is when the consumption good is traded with transaction costs when the CCAPM will not hold.

6.4 CAPM versus CCAPM

This subsection aims to compare the theoretical basis of the CCAPM with that of the traditional Sharpe-Lintner CAPM. From the approach taken in this chapter, we know that there is an abstract link between the two from the fundamental theorem of asset prices. The links are much closer than this, though. First, notice that the discrete time (one period) version of the CCAPM was derived under (i) quadratic utility and complete markets and (ii) bivariate normal asset returns/consumption and complete markets. It is very well known, though, that in a one period complete market model, either quadratic utility or normal asset returns is sufficient to generate the variance-aversion that leads to the Sharpe-Lintner CAPM. In other words, as emphasised by Duffie and Zame (1989), “There’s only one CAPM” (heading to their section 2). This is because aggregate consumption comes from the returns of marketable assets as all assets are marketable. In continuous time

the two become distinguishable. This is because, with state independent utility, the ICAPM and the CCAPM are equivalent. The ICAPM reduces to the CAPM only under the additional assumption that there are no state variables. There is, therefore, a very close link between consumption based and market portfolio based capital asset pricing models.

It is also possible to show that, under certain types of recursive utility functions, an asset's covariance with both market returns and changes in consumption will influence its market price. The debate here is based on that given in Epstein and Zin (1991). Take, for example, the recursive (or Kreps and Porteus (1978)) utility function with the form:

$$\begin{aligned} U_t &= W(c_t, \mu[\tilde{U}_{t+1}|I_t]) \\ W(c, z) &= [(1 - \beta)c^\rho + \beta z^\rho]^{1/\rho} \\ \mu[\tilde{x}] &= [E\tilde{x}^\alpha]^{1/\alpha} \end{aligned}$$

where I_t represents the information available to the agent at time t and the notation is otherwise consistent with that used elsewhere in this section. This utility function degenerates to the simple time-separable isoelastic utility function in the case when $\alpha = \rho$. If $\alpha \neq \rho$ then $1 - \alpha$ can be interpreted as a coefficient of relative risk aversion and $1/(1 - \rho)$ is the elasticity of intertemporal substitution. More generally, this utility function has important characteristics that cannot be captured by time-separable utility³⁸. Epstein and Zin (1991) show that, under this utility function, the Euler equation becomes:

³⁸This topic is too complex to discuss here so the reader is referred to Epstein (1992) and the references therein for a full discourse on such preferences.

$$E_t \left[\beta^\psi \left(\frac{\tilde{c}_{t+1}}{c_t} \right)^{\psi(\rho-1)} \tilde{r}_{mt}^{\psi-1} \tilde{r}_{it} \right] = 1$$

for all i , where $\psi := \alpha/\rho$. Dropping tildes and using upper case letters to denote logarithms (so, for example, $C_{t+1} := \ln(\tilde{c}_{t+1})$), the previous offset equation can be rearranged to give:

$$E_t[\text{Exp}(A + R_{it})] = E_t[\text{Exp}(A + R_{jt})]$$

for any assets i, j and where $A := \psi(\rho-1)\Delta \ln c_t + (\psi-1)R_{mt}$. Next make the assumptions that R_{it}, R_{jt}, R_{mt} and $\Delta \ln c_t$ are normally distributed. These conditions are, of course, not consistent. If r_{it} is lognormally distributed for all i , then r_{mt} , which is a linear combination of the individual r_{it} s will not be lognormally distributed. This point is overlooked in the remainder of the analysis. So, if $A + R_{it}$ is normally distributed, the expectation of the left and right hand sides of the previous offset equation can be calculated. Rearrange to get:

$$E_t[R_{jt} - R_{it}] = \frac{\text{Var}(R_{it}) - \text{Var}(R_{jt})}{2} + (1 - \psi)\text{Cov}(R_{jt} - R_{it}, R_{mt}) - \psi(\rho - 1)\text{Cov}(R_{jt} - R_{it}, \Delta \ln c_t)$$

Therefore asset returns have both a CAPM and CCAPM element. In the case of isoelastic utility where $\alpha = \rho$, $1 - \psi = 0$ and $1 - \rho = 1 - \alpha = \gamma$ this simplifies to give:

$$E_t[R_{jt} - R_{it}] = \frac{\text{Var}(R_{it}) - \text{Var}(R_{jt})}{2} + \gamma \text{Cov}(R_{jt} - R_{it}, \Delta \ln c_t) \quad (26)$$

This, then, is a version of the CCAPM under the assumption of joint lognormality of consumption and asset returns and isoelastic utility. This equation has been used by a number of authors in the empirical tests of the CCAPM.

7 Developing the Mehra & Prescott puzzles

This chapter is concluded by developing the theory of consumption based asset pricing described above into formal expressions for the equity premium and the real riskfree rate. These provide the basis for the Mehra & Prescott puzzles that are described in detail in the next chapter. The key assumptions that underlie all the models in this section are:

- All investors have time-separable power / logarithmic utility with the parameters of time preference and relative risk aversion constant across investors. That is, the preferences of any individual are described in section 4. There are homogeneous expectations.
- Markets are complete. In this case, equilibrium prices are the same as if there were a representative agent whose preferences could also be described by section 4 (see sections 5.1 and 5.2).
- There are no frictions or taxes so that the Euler equation of section 5.3 and the dynamic programming theory of section 6 can be invoked.
- Aggregate consumption is smooth.

The exact distribution assumed for aggregate consumption will vary between models, but smoothness is required throughout. Some models require that the return to the market is also smooth. Theoretical explanations of the Mehra & Prescott puzzles concentrate on relaxing one of these four assumptions. This thesis concentrates on the second assumption — complete markets.

7.1 A one-period economy

7.1.1 Non-parametric

We know, from equation 14, that if consumption is smooth, then the pricing kernel $\pi = U'(c_1, 1)/U'(c_0, 0) \approx \beta(1 - \gamma\Delta\ln c)$. Using equations 4 and 6 for the equity premium and real riskfree rate, it is clear that:

$$E[r_m - r_f] = \frac{\gamma \text{Cov}(r_m, \Delta\ln c)}{1 - \gamma E[\Delta\ln c]} \quad (27)$$

$$r_f = \frac{1}{\beta E(1 - \gamma\Delta\ln c)} \quad (28)$$

It is worth noting that, if $E[\Delta\ln c]$ is small, then this expression for the equity premium is the same as the CCAPM version of the equity premium given in equations 12, 24 and 25.

7.1.2 A parametric approach

It is also possible to develop the equity premium puzzle within a single time period using a parametric approach (see Campbell, Lo and MacKinlay (Forthcoming) and Weil (1994)). Define $v_i := \beta(1 + r_i)(c_1/c_0)^{-\gamma}$. Then

$\ln(v_i) = \ln(\beta) + \ln(1 + r_i) - \gamma\Delta\ln c$. Use $R_i := \ln(1 + r_i)$. Now, if returns and consumption are lognormally distributed, then $\ln(v_i)$ is normally distributed with mean $\ln(\beta) + E[R_i] - \gamma E[\Delta\ln c]$ and variance $\text{Var}(R_i) + \gamma^2 \text{Var}(\Delta\ln c) - 2\gamma \text{Cov}(R_i, \Delta\ln c)$. So, $E[v_i] = \ln(\beta) + E[R_i] - \gamma E[\Delta\ln c] + 0.5\text{Var}(R_i) + 0.5\gamma^2 \text{Var}(\Delta\ln c) - \gamma \text{Cov}(R_i, \Delta\ln c)$. However, from the Euler equation, we know that $E[v_i] = 0$. So:

$$E[R_i] + 0.5\text{Var}(R_i) = -\ln(\beta) + \gamma(E[\Delta\ln c] + \text{Cov}[R_i, \Delta\ln c]) - 0.5\gamma^2 \text{Var}(\Delta\ln c) \quad (29)$$

Substitute in for the riskless asset and the market portfolio to obtain:

$$R_f = -\ln(\beta) + \gamma E[\Delta\ln c] - 0.5\gamma^2 \text{Var}(\Delta\ln c) \quad (30)$$

$$E[R_m - R_f] = \gamma \text{Cov}(R_m, \Delta\ln c) - 0.5\text{Var}(R_m) \quad (31)$$

Notice that equation 31 could also have been derived from equation 26.

7.2 A Merton (1971) style analysis

Consider an infinitely lived economy in the style of that described in section 6 and assume that there are no state variables. As there exists a representative agent with isoelastic utility, the investor is trying to optimise her investment and consumption plan with respect to the objective function $E_0[\int_0^\infty \beta^t c(t)^{(1-\gamma)}/(1-\gamma)dt]$ subject to the usual budget constraints given by equation 15. Let the riskless asset have fixed return r_f and denote the dynamics of the price of the risky asset p_m by $dp_m = p_m\mu dt + p_m\sigma dz$ for an

ordinary Weiner process dz . The key result is given in equation 4.42 of the reprint of Merton (1971) in Merton (1992). It can be shown in this case that the optimal consumption path at any time t is some constant κ times the wealth at that time. Here

$$\kappa = \frac{-\ln(\beta)}{\gamma} - (1 - \gamma) \left[\frac{(\mu - r_f)^2}{2\sigma^2\gamma^2} + \frac{r_f}{\gamma} \right]$$

Notice that because $c_t = \kappa W_t$ then $dc_t/c = dW_t/W$. From the budget condition (equation 15) we have an expression for dW , and hence for dc . As before, w denotes the optimal proportion of wealth in the risky asset:

$$\begin{aligned} dW &= (w(\mu - r_f) + r_f)Wdt + wW\sigma dz - cdt \\ \Rightarrow dc/c &= (w(\mu - r_f) + r_f - \kappa)dt + w\sigma dz \end{aligned}$$

The final step is to identify w . As there are no state variables, it is clear from equation 17 that $w = (\mu - r_f)/\sigma^2\gamma$. Substituting in, we derive an expression for the relative change in optimal consumption given by Constantinides (1990) equations 11,17 and 18 and Ahn (1990)³⁹ equation 11:

$$\frac{dc}{c} = \left[\frac{r_f + \ln(\beta)}{\gamma} + \frac{(\mu - r_f)^2(1 + \gamma)}{2\gamma^2\sigma^2} \right] dt + \frac{\mu - r_f}{\gamma\sigma} dz$$

From this, the predicted equity premium and riskfree rate can be calculated for given mean and standard deviation of changes in consumption.

Using $\Delta \ln c$ to approximate for dc/c :

³⁹There is a slight typographical error in Ahn's equation, which is corrected for here.

$$\begin{aligned} r_f &= -\ln(\beta) + \gamma E[\Delta \ln c] - 0.5\gamma(\gamma + 1)\text{Var}(\Delta \ln c) \\ E[\mu - r_f] &= \gamma \text{Std}(\Delta \ln c)\sigma \end{aligned}$$

7.3 Markov growth model

Consider the standard infinitely lived economy with a representative agent who has power utility. There is one riskless asset in zero net supply and one risky asset. Let income (output) and consumption be constrained to equal the dividends from the risky asset. Denote the value of this variable at time t by c_t . Consumption growth $x_t := c_t/c_{t-1}$ is allowed to take any one of s values, $\{\lambda_1, \dots, \lambda_s\}$ — x_t is the state variable. Note that these growth rates do not depend on t . Consider the transition probability matrix $\phi_{ij} := \text{Prob}(x_{t+1} = \lambda_j | x_t = \lambda_i)$ (again t independent) which describes the evolution in consumption / dividends over time. If consumption at time τ is c_τ and $x_\tau = \lambda_i$, (that is, the economy is in state i) then what is the price $p_m(c_\tau, i)$ of the risky asset? Use $E_i[\cdot]$ to denote expectations conditional on the state at τ being i :

$$\begin{aligned} p_m(c_\tau, i) &= E_i \left[\sum_{t=\tau+1}^{\infty} \beta^t \left(\frac{c_t}{c_\tau} \right)^{-\gamma} c_t \right] \\ &= E_i \left[\sum_{t=\tau+1}^{\infty} \beta^t c_\tau^\gamma (c_t)^{1-\gamma} \right] \\ &= c_\tau E_i \left[\sum_{t=\tau+1}^{\infty} \beta^t (x_{\tau+1} \dots x_t)^{1-\gamma} \right] \\ &= k_i c_\tau \end{aligned}$$

where k_i is a constant that depends only on the state at time τ . Return now to the one period Euler equation:

$$\begin{aligned}
p_m(c_\tau, i) &= \sum_{j=1}^s \beta \phi_{ij} [p_m(c_{\tau+1}, j) + \lambda_j c_\tau] \left(\frac{\lambda_j c_\tau}{c_\tau} \right)^{-\gamma} \\
\Rightarrow k_i c_\tau &= \sum_{j=1}^s \beta \phi_{ij} [k_j \lambda_j c_\tau + \lambda_j c_\tau] (\lambda_j)^{-\gamma} \\
\Rightarrow k_i &= \sum_{j=1}^s \beta \phi_{ij} (k_j + 1) (\lambda_j)^{1-\gamma}
\end{aligned}$$

Given this, it is now clear that r_{mij} , the return to the risky asset between states (growth rates) i and j is given by:

$$r_{mij} = \frac{\lambda_j (k_j + 1)}{k_i} - 1$$

Returning to the one period Euler equation, it is clear that the price of the riskless asset p_{fi} given that the current state (growth rate) is i is given by

$$p_{fi} = \sum_{j=1}^s \beta \phi_{ij} \lambda_j^{-\gamma}$$

7.4 Comparing the models

We have derived four expressions for the equity premium and the riskfree rate that are strongly based on the consumption based asset pricing literature developed earlier in the chapter. The Markov growth model is not easy to present in an easily interpretable form. This section is concluded by presenting the analytical form for the equity premium (EP) and real riskfree rate (RRFR) using non parametric single period, parametric single period and continuous-time respectively. Bear in mind that these three forms are

not directly comparable:

$$\begin{aligned} \text{EP} &= \begin{cases} \frac{\gamma \text{Cov}(r_m, \Delta \ln c)}{1 - \gamma E[\Delta \ln c]} \\ \gamma \text{Cov}(R_m, \Delta \ln c) - 0.5 \text{Var}(R_m) \\ \gamma \text{Std}(\Delta \ln c) \sigma \end{cases} \\ \text{RRFR} &= \begin{cases} \frac{1}{\beta E(1 - \gamma \Delta \ln c)} - 1 \\ -\ln(\beta) + \gamma E[\Delta \ln c] - 0.5 \gamma^2 \text{Var}(\Delta \ln c) \\ -\ln(\beta) + \gamma E[\Delta \ln c] - 0.5 \gamma (\gamma + 1) \text{Var}(\Delta \ln c) \end{cases} \end{aligned} \quad (32)$$

8 Conclusion

This chapter has reviewed the main theoretical literature that has developed consumption based asset pricing theory. This is the foundation for the work that follows in subsequent chapters. It was shown that, provided that payouts to assets do not have infinite variance, the absence of arbitrage opportunities alone is sufficient to guarantee the existence of a (finite variance) pricing kernel. This pricing kernel will only be uniquely defined if the market is complete. Work on the theory of asset pricing can be viewed as attempts to accurately identify the functional form of this linear pricing rule. Within a consumption based paradigm, it was shown that the pricing kernel can be formulated as the ratio of marginal utilities of individual consumption. With complete markets, aggregation results can be invoked that allow asset prices to be modelled within a representative agent paradigm. However, if markets are incomplete, aggregate consumption cannot necessarily replace individual consumption in the models. This issue lies at the very heart of this thesis. As the empirical evidence shows that individual consumption

is more volatile than aggregate consumption, the testable implications of complete and incomplete market models are different.

This chapter also reviews the robustness of the CCAPM to different theoretical developments. Essentially, if a representative agent exists and aggregate consumption is sufficiently smooth then the CCAPM follows. The model will not follow if aggregate (individual) consumption and dividends have jump components in a complete (incomplete) market setting. Explicit forms for the equity premium and real riskfree rate are given for economies that are “consistent” with the CCAPM (that is, complete market, smooth aggregate consumption economies). Chapter 3 examines the ability of these models to explain the observed first moments of asset returns in the US over the past century. It will be concluded that the data is at odds with the theory reviewed so far. Attempts to reconcile the theory and data are discussed in detail in the next chapter.

9 Appendix

9.1 Probability space

The mathematics of the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, with a filtration process $\{\mathcal{F}_t\}$ is widely reported in the mathematics and finance literature and therefore only a brief description is given here. First, the intuitive understanding is:

“Tyche, Goddess of Chance, chooses a point ω of Ω ‘at random’ according to the law \mathcal{P} in that, for B in \mathcal{F} , $\mathcal{P}(B)$ represents the ‘probability’ (in the sense understood by our intuition) that the point ω chosen by Tyche belongs to B ”

Williams (1991) p. 23

Because straight mathematical descriptions are so common, the formal mathematics here is presented in relation to an example and trying to use the minimum amount of mathematical terminology. Consider a two period world. At time 1, a die is tossed. At time 2, the die is tossed again. Ω is the set of possible states of the world at $t = 2$. In this example there are 36 such states (6 possible outcomes for the first coin toss \times 6 for the second coin toss). The elements of Ω are denoted by $\{\omega\}$ which here are given by $\{(1, 1), (1, 2), \dots, (6, 6)\}$, where the first number refers to the outcome of the first toss, the second to the second toss. Now consider any *event* (subset) B of Ω that might be of interest. Such an event might be “The sum of the two rolls is more than 10”. Individuals assign $\mathcal{P}(B)$ at $t = 0$ to the probability that the state of nature $\omega \in \Omega$ at $t = 2$ is such that event B has occurred. That is \mathcal{P} maps B onto $[0, 1]$. For this example $B = \{(6, 5), (5, 6), (6, 6)\} \subset \Omega$. If

investors believe that the die is “fair” (that is, each face has a $1/6$ probability of showing on each toss), then $\mathcal{P}(B) = 3/36$.

Consider the collection of all events to which individuals can assign a probability. This is \mathcal{F} . Certain requirements are made of \mathcal{F} :

- If an individual can assign a probability to something happening, then the individual can also assign a probability to it not happening. That is, if $B \in \mathcal{F}$ then $\Omega \setminus B \equiv \{\omega \in \Omega : \omega \notin B\} \in \mathcal{F}$.
- If an individual can assign a probability to two events, then the individual can also assign a probability to them both happening and to either happening. So, if $B_1, B_2 \in \mathcal{F}$ then $B_1 \cap B_2 \in \mathcal{F}$ and $B_1 \cup B_2 \in \mathcal{F}$.
- Individuals assign a probability 0 to nothing happening and 1 to something happening: $\emptyset, \Omega \in \mathcal{F}$ and $\mathcal{P}(\emptyset) = 0, \mathcal{P}(\Omega) = 1$
- Probabilities are additive for mutually exclusive events. So, if $B_1, B_2 \in \mathcal{F}$ and $B_1 \cap B_2 = \emptyset$ then $\mathcal{P}(B_1 \cup B_2) = \mathcal{P}(B_1) + \mathcal{P}(B_2)$.

The probability space $(\Omega, \mathcal{F}, \mathcal{P})$ is defined. The filtration $\{\mathcal{F}_t\}$ refers to the set of information available at time t . “For example, if some subset A of Ω is an element of ... \mathcal{F}_t , then at time t , intuitively speaking, one “knows” whether the “correct state of the world” is an element of A , that is, whether A is “true” or “false”” (Duffie (1988) p. 131). As it is assumed that no knowledge is ever forgotten once learned, $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $t > s$. For our coin tossing example \mathcal{F}_0 has only two elements: \emptyset is false and Ω is true so

$\mathcal{F}_0 = \{\emptyset, \Omega\}$. At time $t = 1$, if a number 2 comes up on the first throw, then $\{(2, 1), \dots, (2, 6)\}$ is true while $\{(1, 1), \dots, (1, 6)\} \cup \{(3, 1), \dots, (6, 6)\}$ is false. These subsets can be “added” to \mathcal{F}_0 to give \mathcal{F}_1 . At time 2, all uncertainty is resolved so $\mathcal{F}_2 = \mathcal{F}$. Within a T -period model, it is standard to assume that $\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_T = \mathcal{F}$.

We define a *random variable (function)* p that is “ \mathcal{F} measurable”. If p maps $\Omega \rightarrow \mathbb{R}$, then for *any* set $Z \in \mathbb{R}$ the set $p^{-1}(Z) \equiv \{\omega \in \Omega : p(\omega) \in Z\} \in \mathcal{F}$. “Intuitively, p is a random variable if, for any possible outcome, we will know whether p has this outcome from knowing the outcome (true or false) of the events in \mathcal{F} ” (Duffie (1992a) p. 223). A process $p = \{p_t\}$ is adapted to $\{\mathcal{F}_t\}$ if, for each t , p_t is \mathcal{F}_t measurable. If we take p to be the price processes of securities then “the assumption that p_t is adapted to $\{\mathcal{F}_t\}$ simply means that among the information available at time t are the prices then prevailing for all traded securities” (Harrison and Kreps (1979) p. 388).

Finally, we define a martingale and an equivalent martingale measure. A process p is a *martingale* (relative to $\{\mathcal{F}_t\}, \mathcal{P}$) if p_t is adapted to $\{\mathcal{F}_t\}$, has finite expectation at all points and $E[p_t | \mathcal{F}_{t-1}] = p_{t-1}$. An *equivalent martingale measure* to $(\Omega, \mathcal{F}, \mathcal{P})$ is a probability measure Q on (Ω, \mathcal{F}) where⁴⁰:

1. $\forall B \in \mathcal{F}, \mathcal{P}(B) = 0 \iff Q(B) = 0$
2. $\frac{dQ}{d\mathcal{P}} \in L^2(\mathcal{P})$
3. The price processes are martingale with respect to Q over $\{\mathcal{F}_t\}$

⁴⁰ $L^2(\mathcal{P})$ means “is square integrable” and is formally defined in the body of the chapter.

9.2 Vector spaces — Definitions

Definition 1 A vector space X is a set of elements called vectors together with two operations. The first operation is addition which associates with any two vectors $x, y \in X$ a vector $x + y \in X$. The second operation is a scalar multiplication which associates with any vector $x \in X$ and any scalar α a vector αx . The following axioms are assumed to hold for all scalars α, β and $x, y \in X$:

$$\begin{array}{rcl}
 x + y & = & y + x \\
 (x + y) + z & = & x + (y + z) \\
 \exists 0 \in X \text{ such that } & & x + 0 = x \\
 \alpha(x + y) & = & \alpha x + \alpha y \\
 (\alpha + \beta)x & = & \alpha x + \beta x \\
 (\alpha\beta)x & = & \alpha(\beta x) \\
 0x & = & 0 \\
 1x & = & x
 \end{array}$$

Definition 2 A nonempty subset M of a vector space X is called a subspace of X if $\forall x, y \in M$ and scalars α, β , the vector $\alpha x + \beta y \in M$.

Definition 3 We say that a vector space X is the direct sum of two subspaces M, N if every vector $x \in X$ has a unique representation of the form $x = n + m$ for $n \in N, m \in M$. We describe this by the notation $X = M \oplus N$ (where \oplus differs from $+$ in that it implies uniqueness).

Definition 4 A transformation from a vector space X into the space of real (or complex) scalars is said to be a functional on X . A functional $f(\cdot)$ is said to be linear if, $\forall x, y \in X$ and scalars α, β , $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$.

Definition 5 A normed linear vector space is a vector space X on which there is defined a real-valued function which maps each element $x \in X$ into a real number $\|x\|$ called the norm of x . The norm satisfies the following axioms:

1. $\|x\| \geq 0 \quad \forall x \in X. \|x\| = 0 \iff x = 0$
2. $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in X$
3. $\|\alpha x\| = |\alpha| \cdot \|x\|$

Definition 6 A functional is continuous at x_0 in vector space X if for every $\epsilon > 0$ there is a $\delta > 0$ such that $\|x - x_0\| < \delta$ implies that $|f(x) - f(x_0)| < \epsilon$. If f is continuous at every point $x_0 \in X$ then f is continuous.

Definition 7 A point $x \in X$ is said to be a closure point of a set P if, given $\epsilon > 0$, there is a point $p \in P$ satisfying $\|x - p\| < \epsilon$. The collection of all closure points of P is called the closure of P and is denoted by \bar{P} . If $P = \bar{P}$ then the set P is closed.

Definition 8 A sequence $\{x_n\}$ in a normed space is said to be a Cauchy sequence if $\|x_n - x_m\| \rightarrow 0$ as $n, m \rightarrow \infty$; ie, given $\epsilon > 0$, there exists an integer N such that $\|x_n - x_m\| < \epsilon \quad \forall n, m > N$.

Definition 9 A normed linear vector space X is complete if every Cauchy sequence from X has a limit in X . A complete normed linear vector space is called a Banach Space.

Definition 10 A Hilbert space is a Banach space X together with an inner product defined on $X \times X$. Corresponding to each pair of vectors $x, y \in X$, the inner product $(x|y)$ of x and y is a scalar. The inner product satisfies the following axioms:

1. $(x|y) = \overline{(y|x)}$
2. $(x + y|z) = (x|z) + (y|z)$
3. $(\alpha x|y) = \alpha(x|y)$
4. $(x|x) = \|x\|^2$

where the bar on the right hand side of axiom 1 denotes complex conjugation.

Definition 11 Two vectors $x, y \in X$ are said to be orthogonal if $(x|y) = 0$.

Definition 12 Given a subset S of a Hilbert space, the set of all vectors orthogonal to S is called the orthogonal complement of S and is denoted by S^T .

9.3 Riesz Representation Theorem

Before proceeding to the main result, it is necessary to establish a preliminary projection theorem. The proof of this result is laborious but trivial, so the reader is referred to pp.50–51 of Luenberger (1969) for the formal proof. Intuitively, the result is clear, as its name would imply. If you are trying to get to a point $x \in X$ from a plane M in the shortest distance, then you project from the point x onto M orthogonally. This projection is unique.

Result 5 The Classic Projection Theorem. *Let H be a Hilbert space and M a closed subspace of H . Corresponding to any vector $x \in H$, there exists $m_0 \in M$ such that $\|x - m_0\| \leq \|x - m\| \forall m \in M$. Furthermore, a necessary and unique condition that $m_0 \in M$ be the unique minimising vector is that $x - m_0$ be orthogonal to M .*

Notice that the classic projection theorem is equivalent to saying that $H = M \oplus M^T$ for any closed subspace M of H . The main result that is

required for the formal proof of the fundamental theorem of asset pricing is now presented:

Result 6 (Riesz-Fréchet). *If f is a continuous linear functional on a Hilbert space H then there exists a unique $y \in H$ such that $\forall x \in H, f(x) = (x|y)$.*

Proof. Define N as the subset of H for which, $f(n) = 0 \iff n \in N$. That N is a (closed) subspace of H is assured through the continuity and linearity of f . If $N = H$, then the theorem holds with $y = 0$. If $N \neq H$ then, from the classic projection theorem, $H = N \oplus N^\perp$ for non-empty N^\perp . Let $z \in N^\perp$ and scale z so that $f(z) = 1$. Now, consider $x - f(x)z$. This vector must be in N as $f(x - f(x)z) = f(x) - f(x)f(z) = f(x)(1 - f(z)) = 0$ through the assumption that $f(z) = 1$. Since $z \in N^\perp$ and $x - f(z)x \in N$, this implies that $(x - f(x)z|z) = 0$. So $(x|z) = f(x)(z|z)$. So, $f(x) = (x|z)/(z|z)$. Defining $y := z/(z|z)$, it is clear that $f(x) = (x|y)$ and so a y exists as required. We now need to prove uniqueness. Suppose there were $y, y' \in H$ with $y \neq y'$ such that $(x|y) = (x|y') = f(x) \forall x$. Then, $(x|y - y') = 0 \forall x \iff y = y'$. So, the uniqueness has also established.

9.4 Applications of the fundamental theorem

This thesis concentrates on the Euler equation representation of the pricing kernel. That is, the pricing kernel is identified throughout this thesis with the ratio of marginal utilities of consumption. This formulation of the fundamental theorem leads, with a few additional assumptions outlined in the body of the chapter, to the CCAPM. This appendix aims to show the strength and

flexibility of the fundamental theorem by considering alternate forms for the pricing kernel that will give other well known asset pricing models.

9.4.1 The Capital Asset Pricing Model

This discussion is based on that given in Duffie (1988). Consider the subspace of tradable assets M . Assuming that the number of traded assets is finite, M is closed (p.93 of Duffie (1988)) and any closed subspace of a Hilbert space is a Hilbert space under the same inner product (p.132 Bollobás (1990)). The Riesz Representation Theorem can now be applied to M so that the price $p(x)$ of any traded payoff $x \in M$ is given by $p(x) = E^P[x\pi]$ for $\pi \in M$. In other words, π is a traded asset, sometimes called the “pricing asset”.

Suppose that a riskless asset exists. Given that this asset pays 1 in all states, it will be denoted by 1_Ω . Consider the equilibrium choice x^k of an agent k . Given the existence of $1_\Omega, \pi$, we can consider the OLS regression of x^k onto the space spanned by $1_\Omega, \pi$:

$$x^k = A^k + B^k\pi + \epsilon^k$$

where ϵ^k has zero expectations and zero covariance with either 1_Ω or π . Now, the price of k 's portfolio is $p(x^k) = E[x^k\pi] = E[(A^k + B^k\pi + \epsilon^k)\pi] = E[(A^k + B^k\pi)\pi]$ as $E[\epsilon^k] = \text{Cov}[\pi, \epsilon^k] = 0$. In other words, for the same initial cost, the investor could have bought instead the portfolio $y^k = A^k + B^k\pi$, which is an achievable portfolio as both the riskless asset and π can be traded. If our investor is variance-averse, then $y^k \succ x^k$ (where \succ means “is

strictly preferred to”) except in the trivial case when $\text{Var}(\epsilon) = 0$. Therefore, in a mean-variance world, all investors will hold an equilibrium portfolio of the form $x^k = A^k + B^k\pi$. Summing x^k ’s over k it is clear that we end up with a total payout from all portfolios (that is, the market payout, M) of $M = a + b\pi$. Assuming that this payout from the “market portfolio” has non-zero variance so that $b > 0$, it is clear that $\pi = (M - a)/b$. By substituting this in to equation 5 with asset j as the market index, the CAPM follows direct.

9.4.2 Arbitrage Pricing Theory

This section is based on the discussion in Ferson (1995). Take the standard APT assumption that asset returns evolve according to:

$$r_{it} = E_{t-1}[r_{it}] + \sum_{k=1}^K b_{ikt} F_{kt} + u_{it} \quad \forall i$$

where F_{kt} are the K economic factors (with zero expectation) that affect asset values systematically at time t , b_{ikt} is the sensitivity of r_{it} to economic factor F_{kt} and u_{it} is the (zero expectation, zero correlation with the specific factors) idiosyncratic effect at time t . From this assumption about the process generating asset returns it is clear that $\text{Cov}(r_{it}, -\pi_t) = \sum_k b_{ikt} \text{Cov}(F_{kt}, -\pi_t) + \text{Cov}(u_{it}, -\pi_t)$. Substituting into equation 4:

$$E_t[r_{it}] = r_{ft} + \sum_{k=1}^K b_{ikt} \frac{\text{Cov}(F_{kt}, -\pi_t)}{E[\pi_t]} + \frac{\text{Cov}(u_{it}, -\pi_t)}{E[\pi_t]}$$

Now, it is assumed that the residual error terms are diversifiable in a

large economy and so $\text{Cov}(u_{it}, -\pi_t) = 0 \forall i$ ⁴¹. The APT now follows directly by defining the risk premium $\lambda_k := \text{Cov}(F_{kt}, -\pi_t)/E[\pi_t]$. That is, we need to identify the pricing kernel in order to determine the risk premium associated with each of the economic factors.

9.4.3 Binomial Options Pricing

Consider the paper in options pricing theory written by Cox, Ross and Rubinstein (1979). This paper confirmed the Black-Scholes options pricing formula by taking the limiting case of a discrete time binomial share price process. While this model is based on pure no-arbitrage (with no additional assumptions about preferences) and hence is fundamentally “different” from other models in this thesis (apart from the APT that has just been reviewed), the power of the fundamental theorem of asset pricing is that it can be used to price contingent claims as well as fundamental securities.

Consider a share, current price S , that has one period to go to expiry. At the end of this time period, the share can have one of two values $S_1 = (1+u)S$ with probability p and $S_1 = (1+d)S$ with probability $1-p$. Suppose that there is also a riskless bond that will pay $(1+r)$ at time 1 with certainty. Let there be a contingent claim on the risky asset that pays C_u if $S_1 = (1+u)S$

⁴¹This is loosely worded in order to enable us to get to an exact pricing relationship. The APT is, in fact, an approximate pricing relationship as the covariance between the error terms and pricing kernel will not, in general, disappear entirely. Much of the theoretical work that has been done on the arbitrage pricing theory has concentrated on working out economies in which the relationship is exact and on calculating bounds on the imprecision in other economies. Some of this work is directly based on the Hilbert space type analysis that has been outlined here. See, for example, Chamberlain and Rothschild (1983), Chamberlain (1983) for this type of approach to the APT.

and C_d if $S_1 = (1+d)S$. We will assume throughout that there are no interim dividend payments or any incentive to exercise early⁴².

This market is complete. There are two possible states of the world and two linearly independent assets. From section 3.2 it is known that the j^{th} elements of the pricing kernel π is given in such a world by the price of the j^{th} pure security. Denote the prices of the pure securities for states u and d by ϖ_u and ϖ_d respectively. Then by buying $(1+u)S$ units of the pure security for state u and $(1+d)S$ units of the pure security for state d then the cash flow for the share has been replicated. Similarly, if we buy $(1+r)$ units of both pure securities, the cash flow for the bond has been replicated. Given the absence of arbitrage:

$$\begin{aligned} S &= (1+u)S\varpi_u + (1+d)S\varpi_d \\ 1 &= (1+r)(\varpi_u + \varpi_d) \end{aligned}$$

It is now clear that the absence of arbitrage implies that $\varpi_u = q/(1+r)$, $\varpi_d = (1-q)/(1+r)$, where $q := (r-d)/(u-d)$. Therefore $\pi^T = [q/(1+r), (1-q)/(1+r)]$ (taking the upstate as state 1 without loss of generality). Given that $p = E[d\pi]$ for all assets, the price of the contingent claim is uniquely defined by arbitrage by:

$$C = \frac{qC_u + (1-q)C_d}{(1+r)}$$

Notice here that q has an obvious interpretation as a probability measure

⁴²One of the strengths of this technique is that it enables any contingent claim to be priced. For the purposes of this exposition, however, examining these situations is not relevant and therefore excluded for simplification.

(it is easily checked that $q \in [0, 1]$) and is, indeed, a well defined EMM).

Turn now to a T period economy. Assume that the interval $[0, T]$ can be divided into n subperiods. During each time period, the share can either have a return of $1+u$ with probability q or $1+d$ with probability $1-q$. Therefore if our n time periods have j up movements and $(n-j)$ down movements in share price, the price of the share will be $(1+u)^j(1+d)^{n-j}S$ at time T whatever the sequence of up and down movements. We can therefore denote the contingent claim payout at this time by $C_{u^j d^{n-j}}$. The risk neutral probability of being in state $u^j d^{n-j}$ after n time periods is $n!q^j(1-q)^{n-j}/(n-j)!j!$. This risk neutral probability measure can be used to value the contingent claim:

$$C = \frac{1}{(1+r)^n} \sum_{j=0}^n \frac{n!}{(n-j)!j!} q^j (1-q)^{n-j} C_{u^j d^{n-j}}$$

Specify the contingent claim so that $C_{u^j d^{n-j}} = \max[0, (1+u)^j(1+d)^{n-j}S - K]$ — a vanilla European call option. There must be an $a \in (0, n]$ such that $(1+u)^{a-1}(1+d)^{n-a+1}S < K$ and $(1+u)^a(1+d)^{n-a}S \geq K$. So:

$$\begin{aligned} C &= \frac{1}{(1+r)^n} \sum_{j=a}^n n C_j q^j (1-q)^{n-j} [(1+u)^j(1+d)^{n-j}S - K] \\ &= S \Phi \left[a; n, \frac{1+u}{1+r} q \right] - \frac{K}{(1+r)^n} \Phi[a; n, q] \end{aligned}$$

where $\Phi[\cdot]$ is the binomial distribution function and notice that $q(1+u) + (1-q)(1+d) = (1+r)$ as required to make the first $\Phi[\cdot]$ term well defined. This is already similar in shape to the now famous option pricing formula of Black and Scholes and getting the final result is just a simple application of the Central Limit Theorem (see Feller (1968), for example).

9.5 Dynamic programming

In this section, dynamic programming is performed to find the weightings w^k in the investor's optimal portfolio. To simplify the notation, we drop the k superscripts for the appendix:

$$\begin{aligned}
 J(W, s, t) &= \max_{(c, w)} E_t \left[\int_t^T U(c, r) dr + B[W(T), T] \right] \\
 &= \max_{(c, w)} \left[E_t \int_t^{t+h} U(c, r) dr + \int_{t+h}^T U(c, r) dr + B[W(T), T] \right] \\
 &= \max_{(c, w)} E_t \left[\int_t^{t+h} U(c, r) dr + J(W, s, t+h) \right]
 \end{aligned} \tag{33}$$

Where there is a boundary condition that $J(W, s, T) = B[W(T), T]$. Now, let $h \rightarrow 0$ and, by the mean-value theorem $\int_t^{t+h} U(c, r) dr \approx U(c, t)h$. We will now take a Taylor's series expansion of $J(W, s, t+h)$ around t .

$$\begin{aligned}
 J(W, s, t+h) \approx & J(W, s, t) + dW J_W + \sum_i ds_i J_{s_i} + h J_t + \frac{1}{2} d^2 W J_{WW} \\
 & + \frac{1}{2} \sum_i \sum_j ds_i ds_j J_{s_i s_j} + \frac{1}{2} h^2 J_{tt} + dW h J_{Wt} \\
 & + \sum_i ds_i h J_{s_i t} + \sum_i dW ds_i J_{s_i W}
 \end{aligned} \tag{34}$$

Now, taking expectations and ignoring $o(h)$ terms:

$$\begin{aligned}
E_t[dW] &= \sum_{i=1}^n w_i(\alpha_i - r)Wh + (rW - c)h \\
E_t[d^2W] &= \sum_{i=1}^n \sum_{j=1}^n (w_i W)(w_j W)\sigma_{ij}h \\
E_t[ds_i] &= f_i h \\
E_t[ds_i ds_j] &= g_i g_j v_{ij} h \\
E_t[ds_i dW] &= \sum_{j=1}^n (w_j W)g_i \sigma_j \eta_{ij} h
\end{aligned}$$

Substituting back into equation 34, and then in turn, substituting from equation 34 to 33, subtracting $J(W, s, t)$ from both sides and dividing through-out by h :

$$\begin{aligned}
0 = \max_{(c, W)} & \left[U[c, t] + J_t + \sum_{i=1}^n (w_i(\alpha_i - r)W + (rW - c))J_W + \right. \\
& \sum_{i=1}^m f_i J_{s_i} + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_i w_j W^2 J_{WW} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m v_{ij} g_i g_j J_{s_i s_j} \\
& \left. + \sum_{i=1}^m \sum_{j=1}^n \eta_{ij} g_i \sigma_j w_j W J_{s_i W} \right]
\end{aligned}$$

Taking the partial differentials gives the two first order conditions. The first of these is called the “envelope condition” and states that the marginal utility from consuming a pound in equilibrium is the same as the marginal utility from investing the pound. This is very similar to the intuition underpinning the Euler equation.

$$\begin{aligned}
U_c(c, t) &= J_W(W, s, t) \\
0 &= J_W(\alpha_i - r) + J_{WW} \sum_{j=1}^n w_j W \sigma_{ij} + \sum_{j=1}^m J_{s_j W} g_j \sigma_i \eta_{ji} \quad \forall i
\end{aligned}$$

At this point, it is convenient to drop into matrix notation. Let $V_{aa} =$

$[\sigma_{ij}]$, $V_{as} = [\sigma_i g_j \eta_{ij}]$, $J_{sW} = [J_{s,iW}]$, $a = [\alpha_i]$. The second first order condition can be restated as:

$$\begin{aligned} 0 &= (a - r1)J_W + V_{aa}WJ_{WW} + V_{as}J_{sW} \\ \Rightarrow W &= \frac{-J_W}{J_{WW}}V_{aa}^{-1}(a - r1) - V_{aa}^{-1}V_{as}\frac{J_{sW}}{J_{WW}} \end{aligned}$$

To return to body of the text, the k superscripts can be added to this equation as no assumptions of homogeneity have been invoked.

Part III

The Mehra & Prescott puzzles

The Mehra and Prescott puzzles⁴³

Abstract

In a seminal 1985 paper Rajnish Mehra and Edward Prescott demonstrated that the observed equity premium is higher than a standard representative agent model can explain while the average real return to riskfree assets is too low. This chapter examines the theory behind these puzzles and the literature that has attempted to explain them away. Possible theoretical solutions include time non-separable utility and incomplete markets. Persistence of idiosyncratic risk is shown to be the crucial factor determining the success of the latter class of model to resolve Mehra & Prescott's puzzles. The ability of potential explanations to fit the Hansen-Jagannathan bound and explain the second moments of asset returns is also discussed. The final section considers other empirical tests of consumption based asset pricing models. It is concluded that the Mehra & Prescott puzzles have been largely resilient to proposed theoretical explanations.

⁴³A version of this chapter was presented as part of the Financial Options Research Centre Seminar Series, University of Warwick, February 1997. I am grateful to the participants for their useful comments.

10 Introduction

This chapter reviews the puzzles of Mehra & Prescott. The structure of the chapter is as follows:

- Section 11 develops the Mehra & Prescott puzzles. In subsection 11.1 evidence is given on the long term average equity premium and real riskfree rate for the US and the UK. This will justify the use of 6.18% and 0.8% respectively for the observed excess rate of return to the market and real riskfree rate in the US to be explained by consumption based asset pricing models. The Mehra & Prescott puzzles are then explained intuitively using graphical evidence. Subsection 11.2 uses the strong theoretical foundations provided in the previous chapter to formally develop the Mehra & Prescott puzzles. Using all four expressions for the equity premium and real riskfree rate given in section 7 it is shown that 6.18% is “too high” for the equity premium and 0.8% “too low” for the real riskfree rate.
- Section 12 considers explanations of the puzzles that do not depend on market incompleteness. In subsection 12.1 we consider data issues. That is, does the observed average excess rate of return to the market over the last century provide an unbiased estimate of the *ex-ante* equity premium to which asset pricing theory refers. It is emphasised how low the precision is of estimates of the averaged excess return to the market. Further, using US / UK data may introduce biases. Subsection

12.2 considers the form of investors' utility functions. In particular, this section allows for time non-separability in the form of investor preferences. While models allowing for habit persistence and recursive utility functions do fit observed financial market behaviour better than time separable utility functions, problems still remain.

- Section 13 considers the most relevant literature to the issues addressed later in this thesis. Incomplete market explanations of the Mehra & Prescott are considered in this section. The section starts with some preliminary issues. In 13.2, the precautionary savings motive is considered. This shows that the equilibrium rate of return to the riskfree asset will be lower in a world with incomplete insurance than a representative agent model if the third derivative of investor preferences is positive. The economics literature on both the theory and empirical support for precautionary savings is briefly reviewed. Subsection 13.3 considers equilibrium asset prices where there is a risky as well as a riskless asset in the economy. We start with a one period model in the spirit of Mankiw (1986). This model is developed in chapters 5 & 6 and so is of particular relevance to this thesis. In this model income risk is highest when dividend payments are lowest⁴⁴. It is shown in this case that the equity premium and real riskfree rate can be very close to the observed values. We then consider the multiperiod models of Heaton & Lucas. Essentially these models show that persistence of idiosyncratic

⁴⁴This should be compared with the model presented in chapter 4 where income and dividend risks are independent.

income shocks plays a central role in determining predicted equilibrium asset returns. If income shocks are short lived then asset markets can help smooth consumption across investors unless there are severe market frictions. The data they analyse suggests that incomplete market models are unlikely to explain the Mehra & Prescott puzzles. Because the persistence of risk is so important in this type of model, the section is concluded by looking at the savings / consumption decisions of the unemployed and the persistence of unemployment risk.

- Section 14 briefly considers two problems that are closely related to the puzzles of Mehra & Prescott. First, the variance bound of Hansen-Jagannathan is examined. Second we consider the ability of the various models to explain the second moments of asset returns. Unsurprisingly, given that incomplete market / time non-separable utility models have not been able to fully explain the first moment of asset returns, they also cannot fully explain the second moments. It is the low volatility of the riskfree rate combined with the high volatility of equity returns that is at the heart of the problem.
- The chapter is concluded by a very brief examination of other empirical tests of consumption based asset pricing models. There are two reasons for presenting this evidence. First, we will use $\gamma = 3, \beta \in [0.97, 1]$ as the “most likely” estimates of the parameters that should be used in a power utility function throughout this chapter. It is not until this

section that this assumption is justified. Second, it should be realised that the Mehra & Prescott puzzles are by no means the only empirical anomalies that result from the representative agent Euler equation / CCAPM paradigm. This section shows that the low covariance between asset returns and any risky asset results in other well documented discrepancies between the theory and observed asset returns.

Published reviews of the Mehra & Prescott puzzles have started to appear recently. See, for example, Heaton and Lucas (1995) (with comment by Zin (1995)), Kocherlakota (1996) and Siegel and Thaler (1997). Existing reviews tend to concentrate on a small part of the problem. For example, Kocherlakota (1996) focuses on the material covered in section 12 of this chapter. Heaton and Lucas (1995) concentrate on the incomplete market explanations which they have been so instrumental in developing. As emphasised in the previous chapter, the discussion given in this thesis combines the theoretical developments of the Mehra & Prescott puzzles given in a number of different sources. The *contribution* of this chapter is that a wide variety of potential explanations are compared. This is, to the authors knowledge, the most comprehensive review of potential explanations of the puzzles as well as the most comprehensive review of the theory that underlies the puzzles. In particular, the strengths and weaknesses of incomplete market models are clearly demonstrated in this chapter.

11 Developing the Mehra & Prescott puzzles

11.1 Background

In this section the intuition behind the Mehra & Prescott puzzles is discussed. Essentially it shall be argued that, because aggregate consumption is growing so fast and so smoothly, the riskfree rate “should be” higher than the observed value and because the variation of aggregate consumption is so low the equity premium “should be” lower than the observed value. First, evidence on the average observed equity premium and real riskfree rate is presented in table 2. This list is not exhaustive but does demonstrate that there is broad agreement on estimates of the real riskfree rate and equity premium over long intervals in the US. UK evidence is also presented. Table 1 in Dimson and Marsh (1994) suggests that these values are similar to the equity premium and real riskfree rate for many established stock markets:

| Paper | Market | Method | $r_m - r_f$ | r_f | $\sigma(r_m - r_f)$ | $\sigma(r_f)$ |
|---------------------------------|-------------------------------|---------------|------------------------------|-------------------------|----------------------------------|----------------------------------|
| Arithmetic/Simple returns | | | | | | |
| mp85 | US - annual 1889 – 1978 | Arith s.e. | 6.18 (1.76) | 0.80 (0.60) | 16.54 (<i>n.a</i>) | 5.67 (<i>n.a</i>) |
| clm93 | US - annual 1892 – 1987 | GMM s.e. | 6.63 (1.78) | 1.19 (0.81) | 19.02 (1.73) | 5.27 (0.74) |
| sieg92 | US - annual 1871 – 1990 | Arith s.e. | 6.5 (<i>n.a</i>) | 1.8 (<i>n.a</i>) | 19.3 (<i>n.a</i>) | 4.7 (<i>n.a</i>) |
| bc96 annualised | US - monthly 1959 – 1991 | Arith s.e. | 5.02 (<i>n.a</i>) | 1.12 (<i>n.a</i>) | 52.68 (<i>n.a</i>) | 3.27 (<i>n.a</i>) |
| ks90 quarterly | US - quarterly 1929 – 1982 | Arith s.e. | 1.78 (<i>n.a</i>) | −0.05 (<i>n.a</i>) | 12.56 (<i>n.a</i>) | 1.46 (<i>n.a</i>) |
| telm93 annualised | US - monthly 1959 – 1986 | GMM s.e. | <i>n.a</i> (<i>n.a</i>) | 0.99 (0.34) | <i>n.a</i> (<i>n.a</i>) | 3.75 (0.26) |
| jenk | UK - annual 1919 – 1992 | Arith s.e. | 7.28 (<i>n.a</i>) | 2.44 (<i>n.a</i>) | (<i>n.a</i>) (<i>n.a</i>) | (<i>n.a</i>) (<i>n.a</i>) |
| dm94 | UK - annual 1955 – 1993 | Arith s.e. | 9.19 (<i>n.a</i>) | 1.69 (<i>n.a</i>) | (<i>n.a</i>) (<i>n.a</i>) | (<i>n.a</i>) (<i>n.a</i>) |
| Geometric / Logarithmic returns | | | | | | |
| camp | US - annual 1889 – 1990 | Log s.e. | 4.2 (<i>n.a</i>) | 1.8 (<i>n.a</i>) | 18.0 (<i>n.a</i>) | 3.3 (<i>n.a</i>) |
| gms87 | US - annual 1890 – 1980 | Log s.e. | 4.0 (<i>n.a</i>) | 1.8 (<i>n.a</i>) | 17.1 (<i>n.a</i>) | 5.6 (<i>n.a</i>) |
| sieg92 | US - annual 1871 – 1990 | Geom s.e. | 4.8 (<i>n.a</i>) | 1.7 (<i>n.a</i>) | (<i>n.a</i>) (<i>n.a</i>) | (<i>n.a</i>) (<i>n.a</i>) |
| jenk | UK - annual 1919 – 1992 | Geom s.e. | 5.79 (<i>n.a</i>) | 1.49 (<i>n.a</i>) | (<i>n.a</i>) (<i>n.a</i>) | (<i>n.a</i>) (<i>n.a</i>) |
| dm94 | UK - annual 1955 – 1993 | Geom s.e. | 6.05 (<i>n.a</i>) | 1.02 (<i>n.a</i>) | (<i>n.a</i>) (<i>n.a</i>) | (<i>n.a</i>) (<i>n.a</i>) |

Table 2: Evidence on the first and second moments of the equity premium and real riskfree rate. See main text for key to authors.

The key to authors is as follows: “telm93” = Telmer (1993), “mp85” = Mehra and Prescott (1985), “camp” = Campbell et al. (Forthcoming), “clm93” = Cecchetti, Lam and Mark (1993), “ks90” = Kandel and Stambaugh (1990), “sie92” = Siegel (1992), “jenk” = Jenkinson (undated), “bc96” = Bansal and Coleman (1996), “gms87” = Grossman, Mellino and Shiller (1987), “dm94” = Dimson and Marsh (1994). For Siegel (1992), Kandel and Stambaugh (1990) and Grossman et al. (1987) the quoted standard deviation is for the return to equity and not the equity premium. For Jenkinson (undated) the riskfree assets are long dated gilts rather than short dated government securities. It should be noted that returns are quoted in three forms in this table: arithmetic, geometric and logarithmic. The relationship between these three forms is now given. For algebraic simplicity (only) it is assumed that the market pays no dividends. In this case the simple return over any time period $r_{At} := (P_t - P_{t-1})/P_{t-1}$. The logarithmic return is given by $r_{Lt} := \ln(P_t/P_{t-1})$. The arithmetic average of simple returns ($\bar{r}_A := 1/T \sum r_{At}$), the geometric average of simple returns ($\bar{r}_G := \prod [1 + r_{At}]^{1/T} - 1$) and the arithmetic average of log returns ($\bar{r}_L := 1/T \sum r_{Lt}$) are presented in table 2. It will now be demonstrated that $\bar{r}_A - \bar{r}_L \approx \sigma^2(r_L)/2$ and that $\bar{r}_G - \bar{r}_L \approx \bar{r}_L^2/2$. To prove the first of these approximations it will be assumed that r_{Lt} is drawn from a normal distribution with mean \bar{r}_L and variance $\sigma^2(r_L)$ for all t :

$$\begin{aligned}
1 + \bar{r}_A &= \frac{1}{T} \sum_{t=1}^T [1 + r_{At}] \\
&= \frac{1}{T} \sum_{t=1}^T \text{Exp}(r_{Lt}) \\
&= \text{Exp}(\bar{r}_L + 0.5\sigma^2(r_L)) \\
&\approx 1 + \bar{r}_L + 0.5\sigma^2(r_L)
\end{aligned}$$

For the second relationship:

$$\begin{aligned}
\bar{r}_L &= \frac{1}{T} \sum_{t=1}^T \ln \left[\frac{P_t}{P_{t-1}} \right] \\
&= \ln \left[\prod_{t=1}^T \frac{P_t}{P_{t-1}} \right]^{1/T} \\
\Rightarrow \text{Exp}(\bar{r}_L) &= 1 + \bar{r}_G \\
\Rightarrow \bar{r}_L + 0.5\bar{r}_L^2 &\approx \bar{r}_G
\end{aligned}$$

Next consider the relationship between the market and aggregate consumption, which, from the previous chapter, has been demonstrated to be the key variable determining equilibrium asset prices. Look at figure 1, which compares real aggregate consumption on non-durables in the UK with the real total return on the (UK) All-Share index. It is apparent that the market crashes of 1974 and 1987 had virtually no effect on consumption at the aggregate level, suggesting that, within a CCAPM framework, all market risk is non-systematic. As aggregate consumption is so smooth, this makes the predicted risk premium on *all* asset classes using the CCAPM very small. It is clearly going to be extremely difficult to generate an equity premium of 6.18% on the basis of such smooth consumption. The riskfree rate puzzle arises because consumption is growing so fast and so smoothly. If the consumption pattern of each investor is perfectly correlated with aggregate

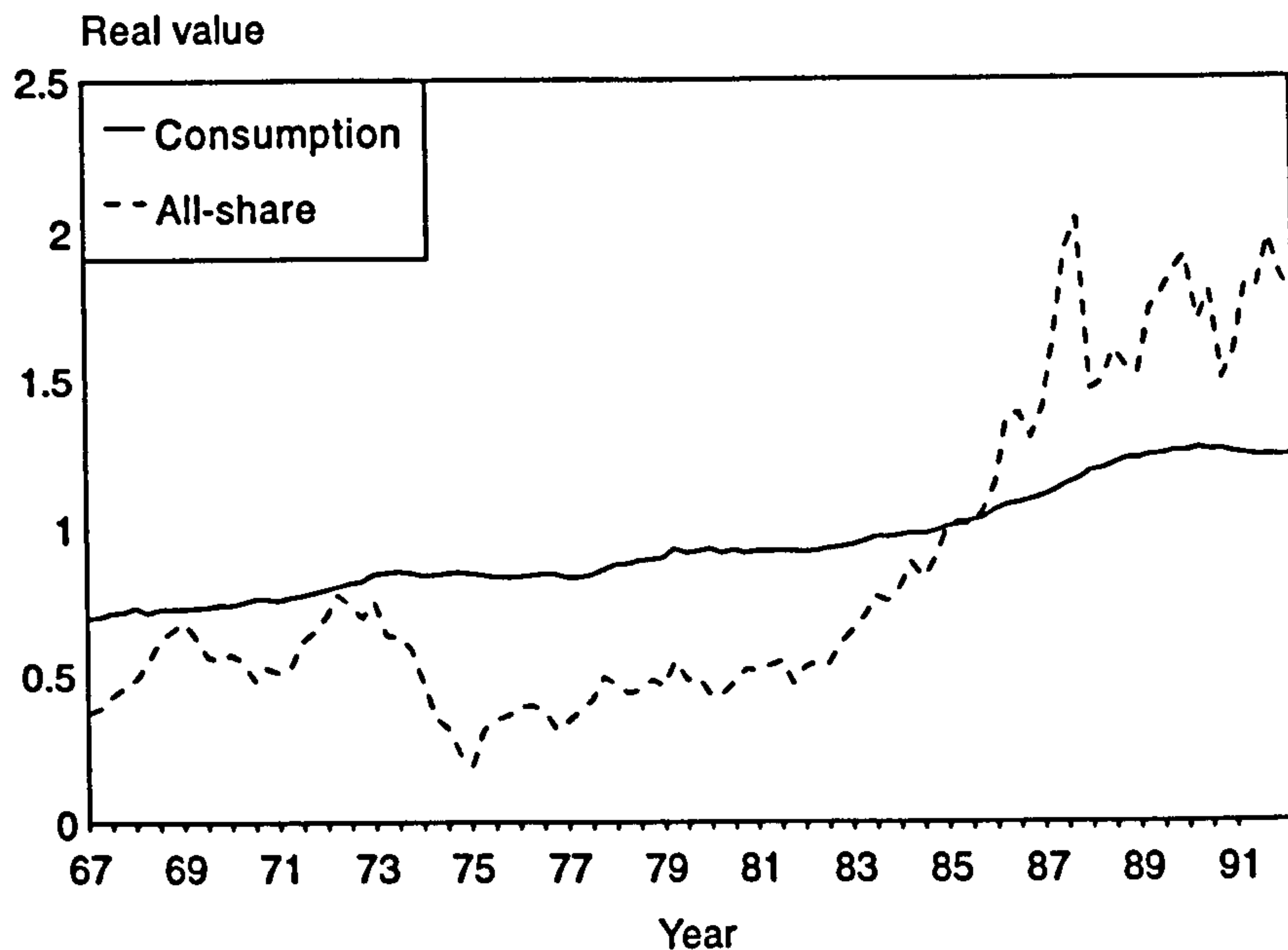


Figure 1: Quarterly real aggregate consumption on non-durables in the UK and the real value of the All-share index with dividends reinvested 1967Q1 – 1992Q2. Both series are readjusted so that the 1985Q1 value is 1.

consumption then it is difficult to see why any investor should wish to defer spending. Only with a high riskfree rate in a standard CCAPM model can equilibrium be maintained. This intuition may help the reader to understand the more formal developments in the next subsection.

11.2 Formal development

The Mehra & Prescott puzzles can be developed in many ways. This section is not exhaustive but does take a number of different approaches to estimating equilibrium asset returns⁴⁵. In the previous chapter, four representations of

⁴⁵A different approach is taken by Hlawitschka and Tucker (1995). They develop the equity premium puzzle from mean-variance analysis. They argue that given three assets (stocks, bonds, riskfree assets) the optimal portfolio would have consisted almost entirely

the real riskfree rate and equity premium were given. Three were presented in equation 32, which is reproduced below:

$$\begin{aligned} \text{EP} &= \begin{cases} \frac{\gamma \text{Cov}(r_m, \Delta \ln c)}{1 - \gamma E[\Delta \ln c]} \\ \gamma \text{Cov}(R_m, \Delta \ln c) - 0.5 \text{Var}(R_m) \\ \gamma \text{Std}(\Delta \ln c) \sigma \end{cases} \\ \text{RRFR} &= \begin{cases} \frac{1}{\beta E(1 - \gamma \Delta \ln c)} - 1 \\ -\ln(\beta) + \gamma E[\Delta \ln c] - 0.5 \gamma^2 \text{Var}(\Delta \ln c) \\ -\ln(\beta) + \gamma E[\Delta \ln c] - 0.5 \gamma (\gamma + 1) \text{Var}(\Delta \ln c) \end{cases} \end{aligned} \quad (35)$$

The fourth representation is given in the Markov growth model of section 7.3. Within any one of these models, there are two equivalent ways of presenting the Mehra & Prescott puzzles. β, γ can be presented exogenously and the equity premium and real riskfree rate inferred. In the second representation we can make β, γ endogenous and have the equity premium and real riskfree rate exogenous.

$$\begin{aligned} \gamma &= \begin{cases} \frac{E[r_m - r_f]}{\frac{\text{Cov}(r_m, \Delta \ln c) + E[\Delta \ln c] E[r_m - r_f]}{E[R_m - R_f] + 0.5 \text{Var}(R_m)}} \\ \frac{\text{Cov}(R_m, \Delta \ln c)}{E[\mu - r]} \\ \frac{\text{Std}(\Delta \ln c) \sigma}{[(1 + r_f)(1 - \gamma E(\Delta \ln c))]^{-1}} \end{cases} \\ \beta &= \begin{cases} \frac{E[\mu - r]}{\text{Std}(\Delta \ln c) \sigma} \\ \frac{[(1 + r_f)(1 - \gamma E(\Delta \ln c))]^{-1}}{\text{Exp}[-R_f + \gamma E[\Delta \ln c] - 0.5 \gamma^2 \text{Var}(\Delta \ln c)]} \\ \frac{E[\mu - r]}{\text{Exp}[-r + \gamma E[\Delta \ln c] - 0.5 \gamma (\gamma + 1) \text{Var}(\Delta \ln c)]} \end{cases} \end{aligned} \quad (36)$$

Remember that r_m, r_f refer to simple arithmetic returns, R_m, R_f refer to logarithmic returns and μ, σ, r refer to the instantaneous returns on the market and riskfree asset in a Merton (1971) style environment. In order to demonstrate the puzzles we need parameters. On the Mehra and Prescott of equity and t-bills since 1970 with virtually no bonds.

(1985) data $E[r_m - r_f] = 0.0618$, $r_f = 0.008$, $E[\Delta \ln c] = 0.0183$, $\text{Std}(\Delta \ln c) = 0.0357$, $\text{Std}(r_m) = 0.1654$ and $\text{Corr}(r_m, \Delta \ln c) = 0.33$. Set $\gamma = 3$, $\beta = 0.98$. If shares are lognormally distributed then $E[R_i] = E[r_i] - 0.5\text{Var}(r_i)$. Using the Mehra and Prescott data this gives $E[R_m] = 0.0481$. Hull (1989) pp.87–8 shows that, over a discrete single period (in this case a year) $\mu = R_m + 0.5\text{Var}(R_m)$. We can therefore assume that $\mu = r_m$ (as done by Ahn (1990), Constantinides (1990)). All other parameters are taken to be the same for each representation of the puzzle. Under these parameters the predicted equity premium is 0.62%, –0.78%, 1.77% for the three models respectively. The predicted real riskfree rate is 7.97%, 6.94%, 6.75% respectively. Alternatively, by setting $\gamma = 20.1, 31.7, 10.47$ and $\beta = 1.57, 0.95, 1.11$, respectively the observed equity premium and real riskfree rate are given. As will be discussed later in the chapter, these estimates of γ are much higher than we might suppose reasonable⁴⁶. For example, if we take an investor with wealth £10,000, and $\gamma = 15$, she would pay 6% of wealth to avoid a 50:50 gamble of \pm £1,000. It is, of course, very difficult to justify $\beta > 1$ as this implies that, *ceteris paribus*, investors prefer to invest later rather than sooner, although Kocherlakota (1990c) has shown that a value of $\beta > 1$ is not necessarily inconsistent with equilibrium in a growing economy.

The puzzles can also be displayed within the Markov growth model as original presented by Mehra & Prescott. Refer to section 7.3 for the relevant notation. In the main parameterisation given by Mehra and Prescott (1985),

⁴⁶See Kandel and Stambaugh (1990) for a defense of $\gamma = 55$, though.

$s = 2, \lambda_1 = 1.054, \lambda_2 = 0.982, \phi_{11} = \phi_{22} = 0.43, \phi_{12} = \phi_{21} = 0.57$. Let $\beta = 0.98, \gamma = 3$. From these:

$$\begin{array}{llll} k_1 & = & 18.636 & k_2 = 18.313 \\ r_{m11} & = & 11.06\% & r_{m12} = 1.77\% \\ r_{m21} & = & 13.01\% & r_{m22} = 3.56\% \\ p_{f1} & = & 0.9498 & p_{f2} = 0.9229 \\ E[r_m - r_f] & = & 0.49\% & E[r_f] = 6.87\% \end{array}$$

Again the Mehra & Prescott puzzles are clear. For $\beta = 1.13, \gamma = 19$ a real riskfree rate of 0.52% and equity premium of 5.54% is predicted. Finally the Mehra & Prescott puzzles are presented in graphical form. The issue is to find parameters β, γ that will generate a sufficiently low real riskfree rate and a sufficiently high equity premium. Consider the three models given in equation 35. If β is restricted to be in the range $[0, 1]$ then the riskfree rate is minimised for any given γ by setting $\beta = 1$ ⁴⁷. So by setting $\beta = 1$ and varying γ , a bound can be generated for the maximum equity premium for any given real riskfree rate. These bounds are presented for the four models in figures 2, 3. The graph that needs the most comment is the top graph in figure 3. This is a variation of figure 1 of Mehra and Prescott (1985). There are certain differences between the figure presented here and the one in the original paper. First, their figure is hand-drawn and therefore somewhat inaccurate. By producing computer output, greater precision on the bound is achieved here, particularly for low $E[r_f]$. Second, they terminate their bound at $E[r_f] = 4\%$ so by going to $E[r_f] = 10\%$ the figure produced here shows

⁴⁷Reducing beta increases the attractiveness of consumption today against consumption tomorrow. So, the lower beta, the less inclined investors are to save. The riskfree rate needs to be increased to maintain equilibrium as beta decreases.

a much wider region. Third, they produce a bound for $\beta \in [0, 1], \gamma \in [0, 10]$ by searching a narrow grid in β, γ space. The figure given here presents 150 “random” points in the narrower space $\beta \in [0.95, 1], \gamma \in [0, 10]$. 75 of these points are for $\beta = 1$ ⁴⁸ and 75 are for $\beta < 1$. γ is chosen at random (rectangular distribution within the admissible) for all 150 points.

12 Possible explanations of the equity premium puzzle

Since the initial observation by Mehra and Prescott (1985) that the average historical riskfree rate was lower in the US than the consumption CAPM using aggregate consumption data, perfect markets and isoelastic utility could explain and that the equity premium was too high, there have been several attempts to explain these “risk-free rate” and “equity premium” puzzles. Duffie (1992b) divides possible reasons of a representative agent model to fail into seven categories: transaction costs, short sales and borrowing constraints, the form of investors utility function, non-stationarity of the relevant stochastic process, failure of “rational expectations behaviour”, noisy data and incomplete markets. Kocherlakota (1996) argues that there can only be three potential explanations to the puzzles: incorrect specification of the utility function, market incompleteness or transactions costs⁴⁹. Here the au-

⁴⁸For the other three models setting $\beta = 1$ gives the bound. This is not proven here in this Markov-growth economy. It is reasonable to expect that, by setting $\beta = 1$, the riskfree rate would be minimised for a given equity premium in this case as well. These 75 points may be expected to describe the bound. The other 75 points informally check this by seeing if, for $\beta < 1$, a point can be generated outside this bound. There is no evidence that this informal bound is violated.

⁴⁹This author believes that this misses at least one important category of potential explanation — that the data used by Mehra & Prescott gives a biased estimate of the

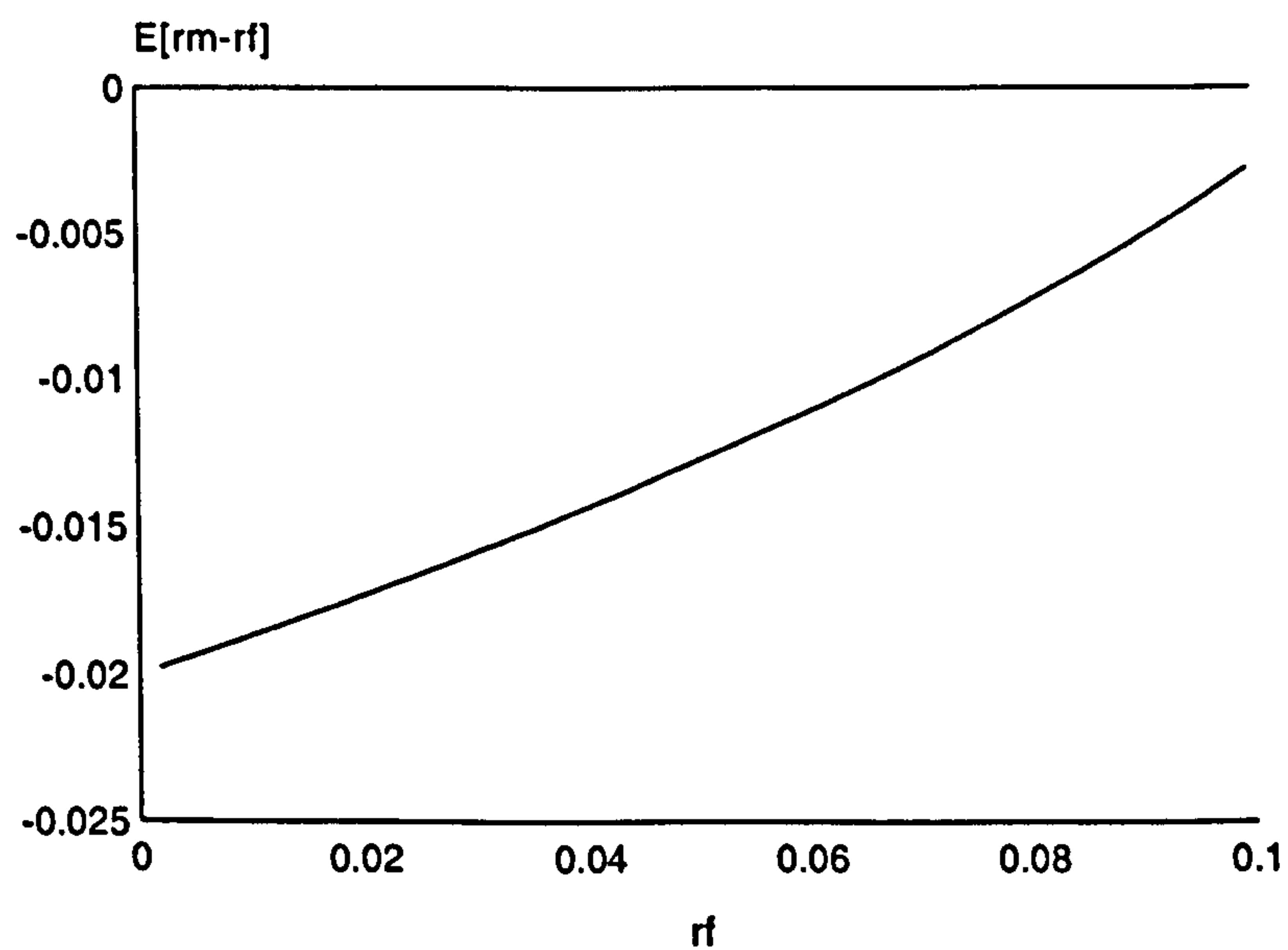
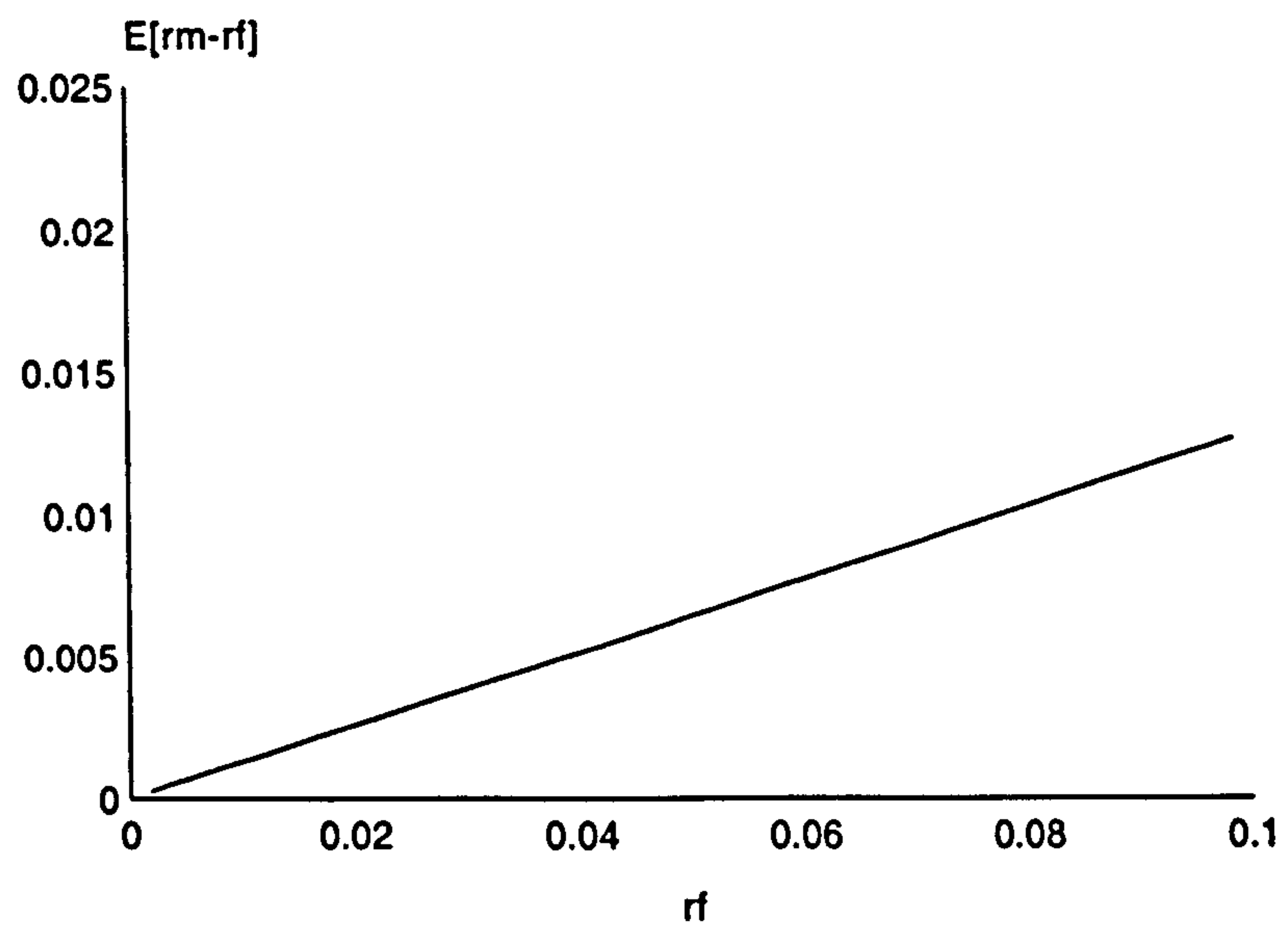


Figure 2: The riskfree rate / equity premium bound (below the line). The top graph is given by a simple one period model where consumption is assumed to be smooth. The bottom graph is given by a one period model in the style of Hansen and Singleton where consumption growth and the return to the market are assumed to be lognormal.

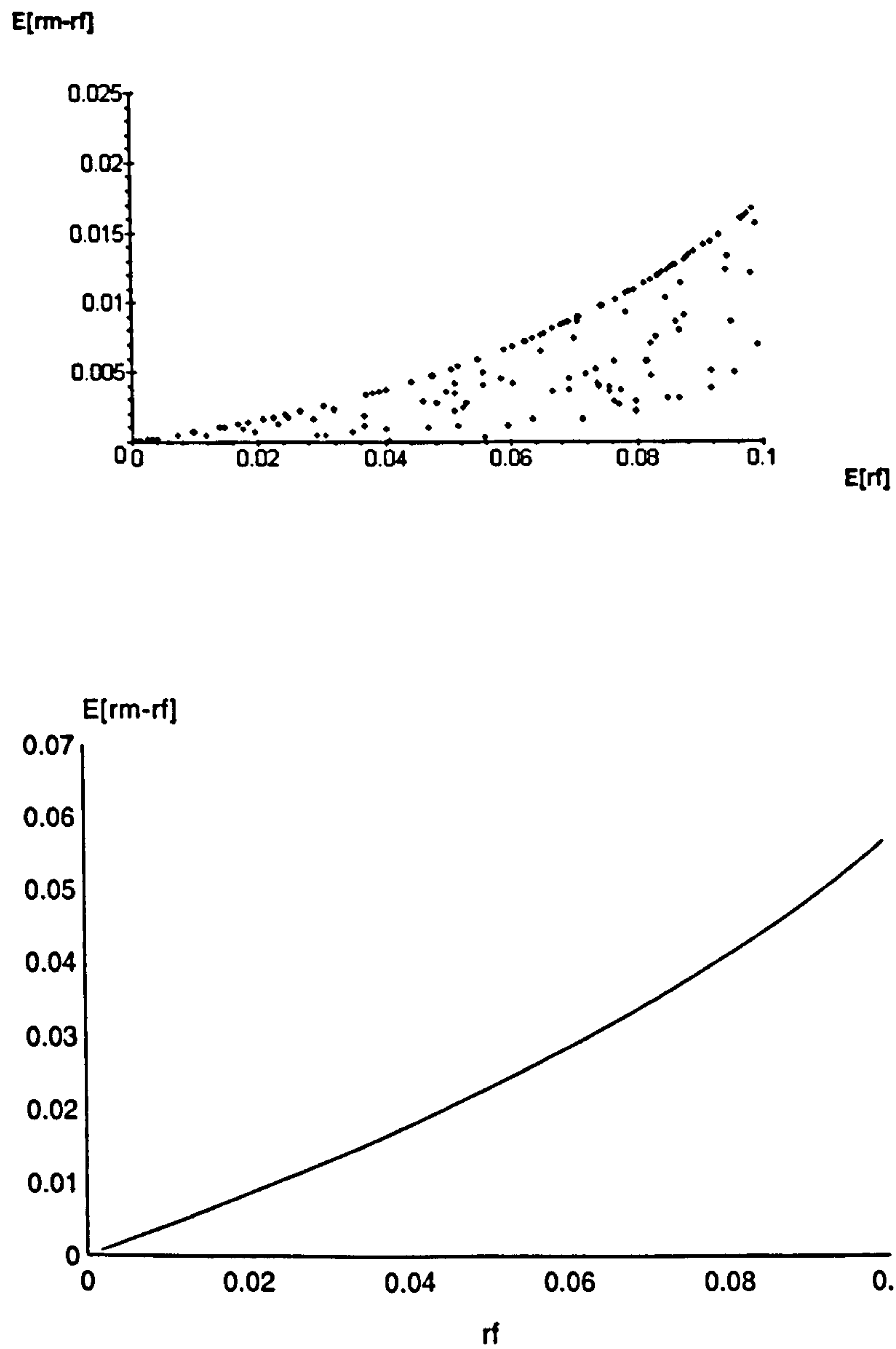


Figure 3: The riskfree rate / equity premium bound. The top graph is given by a Markov-growth model. This figure can be directly compared with figure 1 in Mehra & Prescott (1985). The bottom graph is given by a continuous-time model in the spirit of Merton (1971).

thor divides the possible explanations for the Mehra & Prescott puzzles into slightly different divisions to these:

12.1 Are these really puzzles?

The puzzles of Mehra & Prescott are that the average real return to the riskless asset in the US over the period 1889–1978 of 0.80% and equity premium of 6.18% are difficult to reconcile with standard equilibrium asset pricing models. There is a stream of literature that debates, though, whether these numbers give an unbiased estimate of the *ex-ante* equity premium and real riskfree rate to which asset pricing theory refers. First, note that Mehra and Prescott (1985) quote standard errors associated with these returns of 0.60% and 1.76%. Using a t-statistic of 2 as a bound, Kocherlakota (1996) argues that the equity premium can reject the hypothesis that $\gamma \leq 6.5$ and the riskfree rate can reject the hypothesis that $\gamma \geq 1$ (for $\beta = 0.99$). While this is still a puzzle, it is less of a puzzle than the spot estimates would imply. Figure 1 of Cecchetti et al. (1993) provides 95% confidence intervals for estimates of the average riskfree rate and equity premium. Interpolating from this graph, a riskfree rate of 2.5% real and equity premium of 4% lies within this confidence interval. These values again significantly reduce the puzzles of Mehra & Prescott, particularly when errors in estimating the parameters of consumption growth are also taken into account.

If we consider, for the moment, only the Markov representation of the puzzles, then there are two potential conflicts with the data. First, in the *ex-ante* equity premium and real riskfree rate.

Mehra and Prescott (1985) representation, aggregate consumption is constrained to equal total dividends and total output. Mehra & Prescott use consumption growth, standard deviation of consumption and first order serial correlation of consumption to calibrate their economy. It is not clear whether the correct macroeconomic variable is being used to estimate these parameters. Also, by assuming that output equals dividends, the additional risk taken by equity holders is not reflected. Mehra and Prescott (1985), in an alternate specification, allow for dividends to be distributed after a large “fixed” cost has been allocated to employees and bond holders. Surprisingly, this makes little difference to the predicted equity premium. However Benninga and Protopapadakis (1990) argues that this is because Mehra and Prescott (1985) restrict positive time preferences. For $\beta = 1.114$, $\gamma = 10$ and debt/market value = 60%, Benninga and Protopapadakis (1990) are able to replicate the mean and standard deviation of returns to the market and the riskfree asset.

It has been suggested that these problems with Markov growth models can be overcome by using the more general Markov-regime switching model of Hamilton (1989). Cecchetti, Lam and Mark (1990) argues that this type of model can be calibrated to accurately reflect the statistical properties of observed asset returns and consumption data. The discussion given here is based closely on that in Cecchetti et al. (1993) and, as with that paper, it is assumed in this discussion that there are only two states. The extension to n states follows direct. Let C_t, D_t reflect the log of aggregate consumption

and dividend stream to the market index respectively. Assume that these processes develop according to:

$$\begin{pmatrix} C_t \\ D_t \end{pmatrix} = \begin{pmatrix} C_{t-1} \\ D_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha_0^C \\ \alpha_0^D \end{pmatrix} + \begin{pmatrix} \alpha_1^C \\ \alpha_1^D \end{pmatrix} S_t + \begin{pmatrix} \epsilon_t^C \\ \epsilon_t^D \end{pmatrix}$$

Here $\epsilon_t^C, \epsilon_t^D$ are assumed to be independent and identically distributed (*iid*) normal, zero mean variables with covariance matrix Σ . S_t is a switching variable that can take values 0 or 1. If $S_{t-1} = 1(0)$, then the probability that $S_t = 1(0)$ is given by $p(q)$. Therefore, the drift terms are given by α_0 or $\alpha_0 + \alpha_1$ depending on the state. Notice that this economy is described by nine parameters⁵⁰ (two probability parameters, four drift parameters and three variance/covariance parameters). As observed by Cecchetti et. al. (1993, p.27) “We obtain the Mehra and Prescott endowment process by setting $C = D$, $p = q$ and $\Sigma = 0$ ”. The advantages of using this more general process are that consumption and dividends are no longer constrained to be equal and that noise is introduced into the process.

There are three problems with this type of model. First, the restriction that consumption equals dividends only applies to the Markov representation of the puzzles. The other three representations use market and consumption data direct. Breaking this link is hence unlikely to be the true explanation of the puzzles. Second, while Markov switching allows for noise, Kocherlakota (1996) asserts that the t-statistics that he quotes to reject $\gamma \leq 6.5$ (equity premium) and $\gamma \geq 1$ (riskfree rate) are (asymptotically) valid for

⁵⁰Compared with three (two drift and, one probability) for the Mehra & Prescott representation.

Markov switching processes. Finally, by applying Jensen's inequality, Abel (1994) shows that under Markov switching the predicted riskfree rate will be higher and the predicted equity premium lower than with unconditional expectations. Therefore allowing for Markov switching processes will deepen the puzzles.

Perhaps the most telling contribution to this debate was made by Brown, Goetzmann and Ross (1995). Essentially their argument is that most equity market research is done in the US *because* this market has been so successful. They point out that had one invested in an international equity portfolio at the beginning of the century then one would have been invested in Moscow, Berlin, Warsaw, Buenos Aires and Cairo. They argue that more than half of exchanges that were existent at the start of the century have had major trading disruptions since then. Therefore examining the equity premium of the US, UK and other "major" markets during periods of continuous trading introduces a serious survivorship bias into the sample. Rietz (1988) also wondered whether, *ex-ante*, investors were worried about a low probability, highly severe, "disaster" state that, *ex-post*, has not been observed in US data. Consider the following one period economy. At $t = 0$ all investors are homogeneous and consume c_0 . At $t = 1$, let there be a high state (h) with consumption $c_h := y_h + d_h$ for all investors and let this state occur with probability q ⁵¹. Let there also be a low state (l) with consumption $c_l := y_l + d_l$

⁵¹Rietz (1988) explains his economy in terms of a Markov-switching model. The explanation of the puzzle here is more in keeping with the economy described of Mankiw (1986) and the theoretical developments that are to follow.

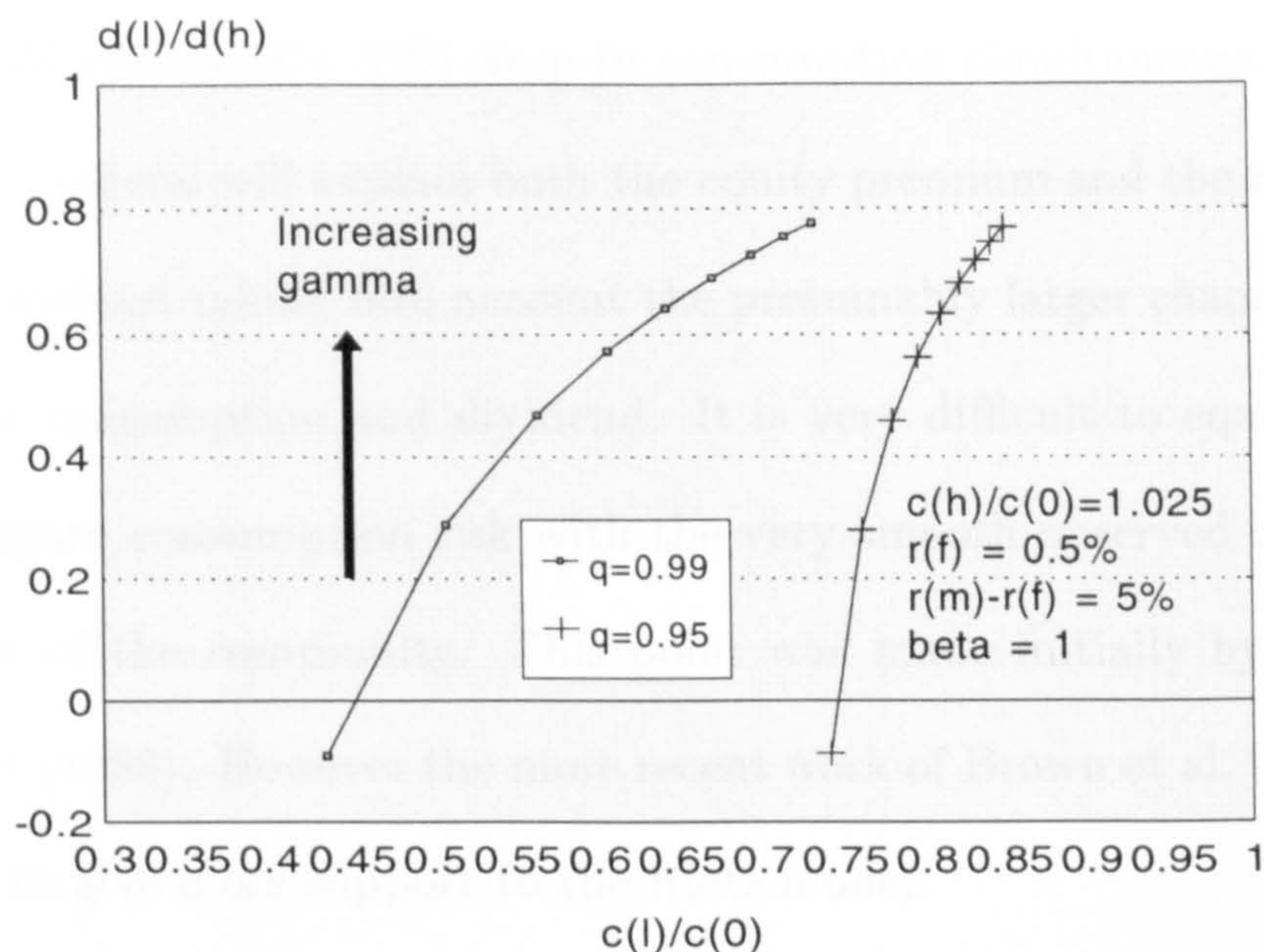


Figure 4: Two state economy, income & dividend risk related, size of risk needed to explain the Mehra & Prescott puzzles.

and probability $1 - q$. Notice that in this case, a drop in dividends necessarily implies a drop in total consumption and hence aggregate shocks are equally shared across the community (thus distinguishing this model from Mankiw (1986)). Denote the expected return to the market by r_m . Then applying the Euler equation with isoelastic utility it is clear that:

$$\frac{c_l}{c_0} = \left\{ \frac{1}{1-q} \left[\frac{1}{(1+r_f)\beta} - q \left(\frac{c_h}{c_0} \right)^{-\gamma} \right] \right\}^{-\gamma^{-1}} \quad (37)$$

$$\frac{d_l}{d_h} = \frac{q \left[\frac{1}{(1+r_m)\beta} - \left(\frac{c_h}{c_0} \right)^{-\gamma} \right]}{\frac{1}{(1+r_f)\beta} - q \left(\frac{c_h}{c_0} \right)^{-\gamma} - \frac{1-q}{(1+r_m)\beta}} \quad (38)$$

Figure 4 demonstrates this relationship for $\gamma \in [2, 10]$ for a set of “re-

alistic" other economic values (see figure). It can be seen that, if $\gamma = 4$, then a 5% chance of a 22% drop in consumption simultaneous with a 54% drop in dividend will explain both the equity premium and the riskfree rate. This is without taking into account the presumably larger chance of smaller drops in consumption and dividend. It is very difficult to equate this size of aggregate consumption risk with the very smooth observed consumption patterns of the community. This point was made initially by Mehra and Prescott (1988). However the more recent work of Brown et al. (1995) would appear to give more support to the Rietz model.

Even ignoring survivorship errors and small-sample problems in estimating averages, Maurice Scott (Scott (1992), Scott (1993)) believes that the spot estimates are flawed. With regard to the riskfree rate, he argues that an *ex-ante* return on a t-bill deflated by an *ex-post* inflation index does not provide a realistic estimate of the real riskfree rate. Instead he believes that an index-linked government bond provides a better estimate. These have been available in the UK since the early eighties and have provided real returns of between 3% and 4% over this period. It is not clear to this author, though, whether this is a true potential explanation to the riskfree rate puzzle. The 1980s were periods of high real returns for all riskfree bonds. Using data from Dimson and Marsh (1994), the real annual rate of return to 3 month treasury bills in London from 1983–1990 varied from 2.8% to 6.9%. Therefore the high return to index linked bonds may be due to the short period under consideration rather than a bias in using t-bills to estimate the true

riskfree rate. Labadie (1989) models an economy with stochastic inflation. While her estimates of the equity premium are higher than in the standard “known” inflation model, the full equity premium puzzle is not explained. Further, including stochastic inflation appears to do little to resolve the risk-free rate puzzle. With regard to the equity premium puzzle, Scott argues that using arithmetic as opposed to geometric rates of return overstate the total return that could be generated from this portfolio. By transferring to geometric rates of return, the equity premium will be around 2.2% lower than the arithmetic average.

Siegel (1992) looks at the long term real return to equity and short term riskless interest rates over a period from 1802. He shows that the long-term average real return to equity is fairly constant over this period but the long-term average real riskfree rate is much lower over the period studied by Mehra & Prescott than over the longer time period. Therefore, if the averages over the two hundred year period of Siegel reflects the true mean required rates of return, the magnitudes of both the riskfree rate and equity premium puzzles are reduced.

So, while spot estimates of the arithmetic average equity premium and riskfree rate in established financial markets provide puzzles for standard representative agent models, these estimates may be biased. Survivorship is clearly an issue as is the inaccuracy of the spot forecast. Further, geometric averages may provide a better (and certainly lower) observed equity premium.

12.2 Form of the investor's utility function

The Mehra & Prescott puzzles comprise of two parts. The equity premium puzzle states that, under isoelastic time-separable utility, \mathcal{R} (coefficient of relative risk aversion) has to be very high to explain the high observed real return to the market. However, the low desire to consumption smooth across time and the low observed real riskfree rate imply that investors have a high elasticity of intertemporal substitution (EIS). For isoelastic utility, where $\mathcal{R} = 1/\text{EIS}$, the riskfree rate puzzle implies a low \mathcal{R} . Therefore, the two puzzles suggest very different estimates of \mathcal{R} . It is a natural assumption, therefore, to use a utility function that has \mathcal{R} and EIS as separate degrees of freedom. There are two main types of utility function that have this form. The first is “habit formation” utility function, which was initially introduced as a potential explanation to the puzzle by Abel (1990) and Constantinides (1990). The second type was briefly reviewed in the previous chapter — Kreps-Porteus utility functions. There is significant literature that examines whether either of these types of utility function can explain away the puzzles. These, though, are not the only type of alternate utility function that has been used to examine the Mehra & Prescott puzzles; for example, Ahn (1989), Ahn (1990) analyses these puzzles with a form of multiplicatively separable utility functions with some success.

“Habit formation” utility functions assume that the utility that an investor derives from this period's consumption is determined by consumption this period relative to consumption in previous periods. That is, if an in-

vestor consumes 50 units today then they will derive a higher utility from this if she consumed 40 units yesterday than 60 units yesterday. Intuitively this form of utility function is justified by saying that if the investor consumed 60 units yesterday then she is “used” to this level of consumption and loses utility from dropping below this level⁵². Such functions have two coefficients (one for risk aversion, one for the degree of habit formation) which separates out \mathcal{R} from the EIS. Essentially habit formation can help to solve the equity premium puzzle as it will induce “stickiness” in an investor’s consumption profile. In times of high income, the investor will not wish to consume all this income because of the potential cost of not being able to match this level of consumption in the future. Similarly, at times of low consumption, investors may wish to borrow against future income to prevent consumption dropping below previous levels. Therefore habit formation implies smooth and upward sloping consumption paths even for reasonable levels of risk aversion, thus providing a potential explanation of for the equity premium puzzle. Unfortunately, though, this is unlikely to be the sole explanation. Because the level of the equity premium is so much higher than the representative agent model can explain the habit formation characteristic must be very strong to resolve the puzzle. Constantinides (1990) shows that, to explain the Mehra & Prescott puzzles, the level of habit formation must be

⁵²Notice, though, that it can also be argued intuitively that if an investor consumes 50 units today then she will derive a *lower* utility from this if she consumed 40 units yesterday than 60 units yesterday. This is justified by saying that if the investor consumed 60 units yesterday then she has “spoiled herself”, or perhaps bought durable goods, and is therefore happy to consume less in the next time period. A utility function that reflects this characteristic is called “durable”. Mathematically, a durable utility function has the same form as a habit formation utility function with a negative coefficient of habit formation.

so strong that it is equivalent to having a subsistence level of consumption at 80% of current consumption— that is, investors are infinitely averse to a drop in consumption of more than 20%. This appears to be inconsistent with observation made in Rietz (1988) (footnote 9) that consumption dropped by 22% in the Great Depression in the US. So, the level of habit persistence required is strongly at odds with additively time-separable isoelastic utility and is highly inconsistent with duration of utility. There are several problems in believing in habit persistence. First, particularly using monthly as opposed to quarterly consumption data, there is more evidence of durability than persistence in preferences (see, for example, Ni (1993) and the references therein). Second, increasing habit persistence implies increasing volatility of IMRS which, in turn, implies higher volatility in the riskfree rate. It is difficult to reconcile this with the smooth observed time series of real riskfree rates (Heaton (1995))⁵³. Third, habit persistence implies positive autocorrelation in consumption growth (see again, for example, Heaton (1995)), which is difficult to reconcile with the low autocorrelation in quarterly consumption and “substantial negative correlation in consumption growth rates at monthly frequencies ”(*ibid.* p.706). So the very high levels of habit formation required to fully explain the Mehra & Prescott puzzles is difficult to reconcile with other observations. Heaton (1995) argues that preferences that exhibit short-term durability with long term persistence are better able to explain market data.

⁵³See subsection 14.2 for a more detailed discussion.

Kreps-Porteus utility functions are also not able to explain the equity premium puzzle. While (quarterly) consumption growth may not be *iid*, to assume that it is so is “... not a blatantly counterfactual assumption” (footnote 13, Weil (1989)). It is commonly believed that, as Kreps-Porteus utility functions have an extra parameter to separate the EIS from \mathcal{R} , they will have more explanatory power than time-separable power utility. Kocherlakota (1990b) argues that when consumption growth is *iid* this is not the case and both models have identical explanatory power. She also examines whether empirical estimates of γ in tests of isoelastic utility should be interpreted as estimates of risk aversion (as is generally the case) or the inverse of elasticity of intertemporal substitution (as Hall (1988) suggests should be the case). Again, under *iid* consumption growth she shows that γ is unambiguously an estimate of risk aversion. Therefore generalising from isoelastic to Kreps-Porteus utility functions will lead to only small changes in estimates of risk aversion as consumption growth only varies marginally from *iid* growth. Therefore, even under Kreps-Porteus utility, an “unreasonably” high coefficient of relative risk aversion is needed to explain the historic excess return to the market. This is clearly demonstrated in table 2 of Kandel and Stambaugh (1991). However, Kreps-Porteus utility can help to explain the riskfree rate puzzle. It is known that, under isoelastic utility, the riskfree rate puzzle and equity premium puzzle can be resolved for values of $\beta > 1$. This is counterintuitive for isoelastic utility as it is believed that investors prefer to consume sooner rather than later. However, under *iid* consumption

growth, if the “true” utility function is Kreps-Porteus but an econometrician estimates the parameters of a power utility function, then the estimate of β is a function of both time preference and EIS (again Kocherlakota (1990a)). So, generalising power utility to Kreps-Porteus utility enables this estimate of $\beta > 1$ observed for isoelastic utility to be decomposed into a true time preference parameter that reflects investors’ desire to consume early and an EIS term that is not necessarily the inverse of the coefficient of relative risk aversion. These points were made initially (and independently from Kocherlakota (1990a)) by Philippe Weil (1989). Epstein and Zin (1990) extend these results to preferences where the coefficient of risk aversion is determined by the standard deviation (as opposed to the usual variance) of the underlying gamble.

So, in general, turning the EIS and \mathcal{R} into separate degrees of freedom within the utility function does not appear to be sufficient to explain the equity premium puzzle. Therefore either the assumption of isoelastic utility is not the driving factor behind the Mehra & Prescott puzzles or economists are yet to specify risk preferences accurately for the representative investor.

13 The Mehra & Prescott puzzles in incomplete markets

13.1 Introduction

This section will review existing literature that is of central relevance for some of the original findings that are to follow in subsequent chapters: the

literature that examines whether nonmarketable risk might help explain the Mehra & Prescott puzzles. This introduction aims to highlight the important distinctions between a world with no nontradable risk, a world where nontradable and marketable risk are independent and a world where the two sources of risk are correlated.

Consider a one period model where there is a risky asset that will pay a dividend \tilde{d} at the end of the period. There is one agent in the economy who will receive exogenous income \tilde{y} at the end of the time period. If equilibrium requires the investor to hold the asset at the beginning of the period and no trading can take place during the period then the period end consumption of the agent \tilde{c} is $\tilde{d} + \tilde{y}$. From the theoretical work on the Euler equation: $E[r_m - r_f] = -\text{Cov}(r_m, \pi)/E[\pi]$ where $\pi = U'(c_1, 1)/U'(c_0, 0)$. While, in most examples to follow, the form imposed on \tilde{d}, \tilde{y} is not necessarily normal, the assumption of normality is now placed on these variables (and hence \tilde{c}) to help the exposition. Under assumptions of normality, the result of Rubinstein (1976) can be invoked that $\text{Cov}[\tilde{d}, U'(\tilde{c})] = E[U''(\tilde{c})]\text{Cov}[\tilde{d}, \tilde{c}]$.

Break this last equation into its two constituent parts. First, from elementary statistics, $\text{Cov}[\tilde{d}, \tilde{c}] = \text{Cov}[\tilde{d}, \tilde{d} + \tilde{y}] = \text{Var}[\tilde{d}] + \text{Cov}[\tilde{d}, \tilde{y}]$. So, for independent marketable and nonmarketable risk, the covariance term is the same as for a world with no income risk. Second, if \tilde{y} is a “small” gamble that is independent to income risk $E[U''(\tilde{c})] \approx E[U''(\tilde{d} + E[\tilde{y}]) + 0.5\text{Var}[\tilde{y}]U''''(\tilde{d} + E[\tilde{y}])]$. This is the first important result (and will be shown below to hold in cases where assumptions of normality are relaxed) that *the impact of nonmar-*

ketable risk to the predicted equilibrium price of an independent marketable risk will be driven by assumptions regarding the first four derivatives (at least) of the utility function. The work on proper risk aversion that is discussed in chapter 4 is driven by considerations about the fourth order derivative of the utility function.

Turning to correlated risks in the style of Mankiw (1986), suppose that dividends at the end of the period can be either high (d_h) with probability p or low (d_l) with probability $1 - p$. If dividends are high then income is high (y_h) with probability 1 . If dividends are low then income will be high with probability q and low (y_l) with probability $1 - q$. In this case $\text{Cov}[\tilde{d}, \tilde{y}] = (1 - p)(1 - q)(E[\tilde{d}] - d_l)(y_h - y_l) = (E[\tilde{d}] - d_l)(y_h - E[\tilde{y}])$. If, in the low state, dividends crash (that is $d_h \gg d_l$) then the equity premium is explained by raising the covariance between dividend and income risk. Crash states in *income* are important for the precautionary savings motive. It has been established that, for an investor with power utility and $\beta = 1$, $r_f = \{E[(c_1/c_0)^{-\gamma}]\}^{-1} - 1$. Consider the following two economies. In the first $c_1 = \{1.04c_0, c_0\}$ with 50:50 probability. In the second $c_1 = \{1.04c_0, 0.64c_0\}$ with 95:5 probability. In both cases $E[c_1] = 1.02c_0$. However, the predicted riskfree rate for $\gamma = 3$ is 5.88% in the first case and - 3.41% in the second. This is a manifestation of the precautionary savings motive. If income is the key source of consumption, it is clear how crash states in income might potentially explain the riskfree rate puzzle. Therefore, by combining the dividend and income crashes, the Mehra & Prescott puzzles might be resolved.

13.2 Precautionary savings and liquidity constraints

This subsection concentrates on the riskfree rate puzzle. The models here cannot be considered to be directly addressing the Mehra and Prescott puzzles as there is only one financial asset. Nevertheless, this section introduces some key ideas that will be extended to tackle the Mehra and Prescott puzzles in the next section. We consider two classes of model that have been used to help explain this violation — the precautionary savings motive and the introduction of borrowing constraints⁵⁴. There is a key difference between most studies that examine precautionary savings and other models reviewed in this thesis. It has been assumed throughout this thesis that aggregate consumption is exogenous and the riskfree rate endogenous. Most studies of precautionary saving specify the rate of return to financial assets and infer from these the savings and consumption decisions of investors.

The concept of precautionary savings motive was introduced by Leland (1968), Sandmo (1970) and Drèze and Modigliani (1972) and this work has recently been updated by Kimball (1990). The Euler equation tells us that the riskfree rate is given by : $1 + r_{ft} = U'(c_t, t) / E[U'(c_{t-1}, t-1)]$. Remember that this relationship holds for the consumption and utility function of all investors. From Jensen's inequality $E[U'(c_t)] \geq U'(E[c_t])$ if and only if $U''' > 0$ with the inequality holding strictly if c_t is stochastic. If the third derivative

⁵⁴This literature is also used to help explain away other apparent violations of the implications of the life cycle /permanent income hypothesis such as the excess sensitivity of consumption to current income and the apparent over-saving of the retired. A review of such general consumption anomalies is outside the scope of this thesis and the reader is referred to one of the excellent published reviews (for example, chapter 6 of Deaton (1992)) for more detailed coverage.

of the utility function is positive, the greater the uncertainty of next period consumption for any investor, the higher their expected marginal utility of next period consumption. If the riskfree rate is exogenous then savings must rise. Such saving is called “precautionary saving”. Kimball (1990) showed that an individual’s (absolute) prudence $\mathcal{P} := -U'''(c)/U''(c)$ measures the strength of an investors’ precautionary savings motive in the way that $\mathcal{A} := -U''(c)/U'(c)$ evaluates her risk aversion. This measure of prudence will be used extensively in chapter 4 and, to a lesser extent, in chapter 5.

Is it reasonable to suppose that investors have a positive third derivative of utility? It can easily be verified that constant or decreasing absolute risk aversion imply a positive third derivative. So, a negative third derivative follows only in certain cases of increasing absolute risk aversion — a condition that does not seem economically “reasonable”. Therefore the conclusion that investors should wish to save at an increasing rate as future income becomes increasingly uncertain is consistent with large classes of preference that financial economists believe to be realistic. For power utility the precautionary savings motive is captured in the terms in $\text{Var}(\Delta \ln c)$ in the bottom two expressions for the real riskfree rate in equation 35. Essentially, as γ grows so terms in $\text{Var}(\Delta \ln c)$ grow increasingly important. “Thus as γ increases, the sign of the relationship between r_f and γ in the time-additive case switches from positive to negative as the precautionary-savings motive overtakes the effect of positive expected consumption growth” (Kandel and Stambaugh (1991) p.56, notation changed).

Another wedge can be driven between the implications of the standard representative agent model and true consumption and savings decisions by introducing risky labour income and a borrowing constraint⁵⁵. This is because, at the time of a low income shock, individuals may not be able to optimally smooth away this shock. Even with quadratic utility (and so no precautionary saving), risky labour income and borrowing constraints, consumption decisions vary substantially from what is predicted by the permanent income hypothesis (see footnote 11 in Zeldes (1989b) and Aiyagari (1994)). Nevertheless, the most “interesting” models will come when there are both borrowing constraints and a precautionary savings motive. This is because, under precautionary saving, one is highly averse to low income states. If borrowing is restricted then it is necessary to save substantially to protect against low income states. An alternative to putting in explicit borrowing constraints is to put in very low endowment states, which will cause investors to have very low borrowing out of choice. Within the papers given below, where utility functions are often assumed to be CRRA (and hence $U''' > 0$) and there is a borrowing constraint, it is often difficult to separate out the two effects.

There are two main ways in which this literature has proceeded. First, one can try to predict the magnitude of the savings that an investor will hold for given income processes with certain borrowing constraints. It is not possible, in general, to get an exact analytical solution and therefore

⁵⁵See Zeldes (1989a) for a test that appears to show that some investors are indeed subject to borrowing constraints.

either approximations are made or numerical techniques are employed — we will call these the “theoretical approaches”. The second technique is to take cross-sectional survey data and look at individual savings and consumption decisions. We will call this the “empirical approach”.

13.2.1 The theoretical approaches

Skinner (1988) takes a second-order Taylor’s series approximation of the Euler equation under the assumption of power utility functions and assumes that both the return on the one financial asset and income are risky (but independent). Because the Taylor’s series is truncated at the second term, it is the variance in the income process that drives the precautionary savings model. By approximating the Euler equation in this way he is able to get an analytical solution to the problem of the predicted level of precautionary savings. As is shown repeatedly in the theoretical models, the level of precautionary saving predicted depends on the level of persistency in the shocks in the income process. Using two parameterisations of the income process Skinner estimates that between 12%–56% of aggregate savings are caused by precautionary saving. Skinner’s model has no borrowing constraints.

By concentrating on exponential utility functions, Ricardo Caballero (Caballero (1990), Caballero (1991)) is also able to derive closed-form solutions to problems involving the precautionary savings motive⁵⁶. Under the assumption that income is a random walk and under plausible parameter estimates,

⁵⁶As he imposes no liquidity constraints on his investors his models allow for negative consumption — a problem that is avoided by power utility functions.

he shows that high consumption growth is consistent with a low riskfree rate. Further, he predicts that substantial savings will be accumulated by individuals under the precautionary motive. If innovations in income are allowed to have large negative skewness, as a proxy for unemployment, then this significantly increases the importance of the precautionary savings motive. This should not, perhaps, be surprising. As income in this model is a random walk and the model specification does not include low probability, high income growth events, unemployment is essentially perpetual in this model. The threat of potential long term unemployment always has a significant impact on individual behaviour in this type of economy.

Zeldes (1989b) uses numerical techniques to estimate the predicted magnitude of precautionary savings in an economy where preferences are represented by power utility functions, there is uncertainty in labour income and no borrowing constraints. Zeldes models two income processes, one with permanent and transitory shocks to income, the other with purely transitory shocks to income and (as we shall see below, somewhat unusually) finds that the predicted level of savings is largely independent of the income process used. His model predicts that investors will hold substantial precautionary savings: on his main parameterisation ($\gamma = 3$, starting wealth = two year's expected income), consumption would be 20% higher than observed levels if investors had no precautionary savings motive.

Turn now to buffer savings. Rearrange equation 30 to give:

$$E[\Delta \ln c] = \gamma^{-1}[R_f + \ln(\beta)] + 0.5\gamma\sigma^2(\Delta \ln c) \quad (39)$$

An investor will be called “impatient” if the growth rate in permanent income g is such that $g > \gamma^{-1}[R_f + \ln(\beta)]$. In this case, and in the absence of income uncertainty and borrowing constraints, the investor would wish to consume soon and repay later. The introduction of the precautionary savings motive (and, possibly, liquidity constraints) will lead the investor to temper the desire to consume early and build up “buffer stock”. The analysis of such behaviour is examined by Deaton (1991), Carroll (1992), Carroll (1997) and Carroll (1996) (see also chapter 6 of Deaton (1992)). The model of Zeldes (1989b) reviewed above, where $g = 0, \beta = 1, R_f = 0$, is the limiting case. Deaton (1991) considers an infinitely lived economy with borrowing constraints. For income shocks that are transitory, he shows that, provided income + wealth is above a certain level, some precautionary saving will take place and a buffer stock will be built. If income + wealth is below a certain target, the investor will find it optimal to completely dissave and then consume hand-to-mouth from income until saving becomes viable again. Consumption is smoothed by precautionary savings in this model. However, the more persistent the income risk, the less effective precautionary savings becomes. This is because the investor pays a high penalty for saving as they are impatient. The more persistent the shock, the longer they will have to save for at times of high wealth. This makes holdings of buffer stock more “lumpy” and does not smooth consumption as successfully in the transitory

shock economy.

Carroll (1992) and Carroll (1997), in economies where there are no explicit borrowing constraints⁵⁷ predict that the variance of consumption growth is inversely related to wealth as poor individuals are less able than the rich to smooth consumption. Therefore, the planned growth in consumption of poor consumers is higher than the planned growth in consumption of rich consumers. So poor consumers are trying to save, while rich consumers are happy to consume wealth (which they wish to do as they are impatient). There exists a target stock level where investors neither wish to save or dissave. Carroll (1997) shows that, at this target level, the growth rate of consumption is marginally below g . Importantly, income uncertainty raises the variance of consumption at a given level of wealth. As planned consumption growth is monotonic decreasing in wealth and as planned consumption growth must be just below g at the target stock level, this means that investors increase their target stock level *ceteris paribus* as income uncertainty increases. This provides Carroll (1992) with a basis to investigate the theory of buffer stocks from a perspective that is of great relevance to this thesis. In a simulated economy he shows that the amount of buffer stock that an investor will wish to hold is highly sensitive to the probability of a transitory (one year) crash to zero income. In his model, the level of savings is predicted to be more than twice as great when the probability of a "zero-income event" is 1%

⁵⁷The models of Christopher Carroll and Zeldes (1989b) include the positive probability of zero income in any time period and hence a positive probability of a consecutive run of zero income events of any finite length. As the marginal utility of consumption at zero is infinite, the introduction of zero income events acts as an "informal" liquidity constraint.

against 0.1%. The impact of such unlikely events on savings levels is much greater than adjusting the standard deviation of transitory and permanent “diffusion” shocks to income.

Some investigations of the role of precautionary savings have been done within a life-cycle model. Hubbard, Skinner and Zeldes (1994) (with comment by Kimball (1994)) have a “standard” precautionary savings model while Carroll (1997) and Carroll and Samwick (1996) have a buffer stock model. The key difference is that, in the former case, savings for retirement happens throughout life. In the latter case holdings of wealth between the ages of 25–50 is almost entirely driven by buffer stock behaviour as individuals are too impatient to save for retirement before then. After 50 there is a predicted sharp rise in savings as people start to prepare for their old age. The two types of model are able to explain different aspects of the age-savings pattern (the former is better able to explain levels of aggregate saving, the latter able to explain the low changes in wealth with changes in the volatility of permanent income) but both perform better than standard life-cycle models.

Huggett (1993) considers a precautionary savings model that has much in common with the models of Heaton and Lucas that will be reviewed below. Nonmarketable risk is associated with unemployment risk when endowment is at 10% of the employed level. It is a multiperiod model with transition probabilities between employed and unemployed states determined by a Markov process. Borrowing constraints are introduced so that individuals have a

strong precautionary savings incentive. The only financial asset is a riskfree asset. For a coefficient of relative risk aversion equal to 3 and a borrowing constraint equal to one year of employed income the predicted riskfree rate is 1.8%. While this is well below the predicted riskfree rate in a representative agent economy, the rate is still higher than the observed value, which is surprising given the shock of unemployment. However, as the average period of unemployment is 17 weeks, the shocks are too transitory to have significant impact on the equilibrium riskfree rate. Notice, though, that the predicted riskfree rate is very sensitive to the borrowing constraint. If this is reduced to four month's employed income then the predicted real riskfree rate is -23%! Aiyagari (1994) considers a model with no aggregate uncertainty but where there is individual uncertainty. Individual endowment shocks are smoother than an unemployment model might suggest, but the persistence and variability in income shocks is modelled on real data⁵⁸. Borrowing is prohibited. The precautionary savings motive / borrowing constraint have only a small effect on aggregate savings and the equilibrium riskfree rate. Only by increasing the size and persistence of idiosyncratic shock to apparently "unrealistic" levels can the riskfree rate puzzle be resolved.

13.2.2 Empirical evidence

The problems with explicit empirical tests of the precautionary savings motive are well summed up by Dardanoni (1991) "The main difficulty in im-

⁵⁸These parameters are very similar to those used by Heaton & Lucas — see below.

plementing the estimation ... is the unobservable nature of almost all the variables involved" (*ibid.* p.156). Despite this, using 1984 UK data he estimates that as much as 60% of all savings may be driven by precautionary motives. Dynan (1993) runs cross-sectional regressions of the growth in log consumption against certain variables including the variance in growth of log consumption. She interprets the coefficient on this variable as being a measure of prudence, as suggested by equation 39. She finds that, by this measure, the precautionary savings motive is small and often indistinguishable from zero. However, as emphasised by Carroll (1997), within a buffer-stock savings model, predicted growth in consumption is approximately equal to the growth in permanent income. That is, prudence is better measured by measuring the size of buffer stocks than looking at growth rates in consumption.

If precautionary motives is an important reason for saving, then people in less secure jobs might be expected to save more than those in safe jobs⁵⁹. The evidence on this is mixed. Skinner (1988), using the Consumer Expenditure Survey of 1972–3 concludes "The savings rates of the self-employed and sales workers, those generally thought to receive riskier incomes, are *less* than the benchmark group of craftsmen" (*ibid.* p. 250, his emphasis). More recent work by Carroll and Samwick (1996), using data from 1981 to 1987 from the Panel Study of Income Dynamics, strongly refutes this. They show that the

⁵⁹There is, though, no general reason why this should be so. People in different jobs may well have systematically different preferences. The comparative savings rates of different occupations may just inform us about these preference differences and tell us little about precautionary savings.

variances of the permanent and transitory components of household income are positively correlated with household savings.

Examining savings rates of investors in the *Surveys of Consumers*, Carroll (1992) demonstrates that individual savings rates are likely to increase at times of high unemployment and times when investors are worried about becoming unemployed. Expected changes in household income, however, has no statistically significant effect on the savings rate. "... consumers both express a desire to save and actually save more when they believe that the unemployment rate will be rising. They also save more when the unemployment rate is high." (*ibid.* p. 105).

Carroll (1992) examines savings ratios in relation to expectations about future unemployment as determined by questionnaire. Kantor and Fishback (1996) examine the issue in a different way. They argue that the level of buffer stocks will be inversely related to income level in a low income state. Therefore, as social insurance increases, so the level of buffer stocks should decrease. They examine saving levels over the 1917–9 period for the US — a time when accident insurance was being widely introduced in the US — and reveal a sharp drop in savings over the period. This is consistent with precautionary savings. Guiso, Jappelli and Terlizzese (1992) approach the problem in a different way. They include two questions on the Italian Survey of Household Income and Wealth survey to assess expectations about future income and inflation uncertainty. They discover that individuals consider their income risk over 12 months to be much lower than the levels usually

assumed in the simulation studies. Therefore, they argue, while precautionary savings might partially account for some of the consumption paradoxes, it is unlikely to be the only explanation.

This concludes the section on precautionary savings where rates of returns on assets are exogenous (and, usually, where only a riskfree asset is available). We turn now to more finance style models where aggregate consumption is exogenous and where a risky asset as well as a riskfree asset exists. The rates of return are endogenous within these models. It will be shown that there are many links between these two streams of literature. Of particular relevance, that low probability, high impact endowment shocks can have significant effects on equilibrium asset prices, particularly if these shocks are persistent, is shown below.

13.3 Risky and riskless assets

We start this subsection with a brief discussion on the implications of having a subset of investors who do not, or will not, invest in financial markets. This is clearly a violation of the representative agent assumptions. Weil (1992b) assumes that the world is divided into those who invest in financial markets and those who do not⁶⁰. In this case, the representative agent for financial asset pricing will appear as some combination of the subset of individuals who participate in financial markets rather than the subset of all individuals (see the discussion in the previous chapter on Brown (1988)). In this case

⁶⁰Weil suggests that such individuals might exist either because they wish to consume hand-to-mouth or because they face infinite transaction costs in financial markets.

using aggregate consumption data in asset pricing models will be misleading. He provides theoretical conditions where the Mehra & Prescott puzzles will be overstated using aggregate data in the presence of non-participants in financial markets. In another paper, Philippe Weil (1994) considers the theoretical implications of the fact that the subset of individuals who participate in financial markets are, in general, more wealthy and have more non-marketable income than normal. With decreasing absolute risk aversion and independence of marketable and non-tradable income, the predicted equity premium is much lower than in the representative agent case. Only with highly correlated labour and financial income can the equity premium puzzle be resolved for the "marginal" investor. Mankiw and Zeldes (1991) take an empirical approach to this problem by examining the consumption profile from individuals who invest in the stock market. They find that these individuals have consumption that is both "more volatile and more highly correlated with the stock market" (pp.98-9) than aggregate consumption data. They argue that a coefficient of relative risk aversion of 6 for investors in the stock market may be consistent with the observed historic excess return to the market. This result should be tempered by noting the poor quality data available to Mankiw and Zeldes (1991). From here on, it will be assumed that all investors have access to, and trade rationally in, financial markets.

13.3.1 A one period model

It has been well established elsewhere in this thesis that if markets are incomplete then the volatility of individual consumption will be higher than the volatility of aggregate consumption. Kahn (1990) shows, though, that both the theoretical and empirical evidence point to a low standard deviation of idiosyncratic endowment risk. “If idiosyncratic risk alone is to account for the equity premium, it is necessary that agents (1) believe that they face a small possibility of severe (greater than 95 percent) drops in consumption and (2) believe that such a possibility is made considerably worse (or more likely) by the holding of equity” (*ibid.* p.42). This subsection follows the spirit of Mankiw (1986) and model 1 of Heaton and Lucas (1992). At $t = 0$ consumption for all *ex-ante* homogeneous investors is c_0 . At $t = 1$ the economy is in an upstate with average per-capita consumption c_h with probability q or in a low state with probability $1 - q$ and average per-capita consumption of c_l . However, in the low state, a proportion $1 - \lambda$ retain consumption at c_h ⁶¹. Therefore, the subset λ of the population where the drop in consumption it is concentrated, ex-post, has consumption $c^* := (c_l - (1 - \lambda)c_h)/\lambda$. So:

$$\tilde{c}_k(1), \tilde{d} = \begin{cases} c_h, d_h & q \\ c_h, d_l & (1 - q)(1 - \lambda) \\ c^* = (c_l - (1 - \lambda)c_h)/\lambda, d_l & (1 - q)\lambda \end{cases}$$

In this case:

⁶¹Notice that this implies that the labour income for the subset of the population that retains its level of consumption must *rise*, as the income that they receive from dividend drop.

$$\frac{c_l}{c_0} = \lambda \left\{ \frac{1}{(1-q)\lambda} \left[\frac{1}{(1+r_f)\beta} - [q + (1-q)(1-\lambda)] \left(\frac{c_h}{c_0} \right)^{-\gamma} \right] \right\}^{-\gamma^{-1}} + (1-\lambda) \frac{c_h}{c_0}$$

d_l/d_h is the same as in equation 38. This is an important discovery that has not been highlighted (to the author's knowledge) elsewhere in the literature. Aggregate crash models of the type suggested by Rietz (1988) are, for a given r_f , equally able to explain the equity premium puzzle as the concentrate shock model of Mankiw (1986). Where the latter class of model does better than the former class is in (i) explaining the riskfree rate puzzle through an increased precautionary savings motive and (ii) appearing to be more economically realistic. The process for $c_l/c_0, c^*/c_0$ that will explain the riskfree rate are shown in figure 5. As can be seen, if the shock is concentrated in a small subset of the population, aggregate consumption hardly needs to fall at all. However, the drop in consumption for the subset of the population where the risk is centred is very severe indeed. Suppose there a 5% chance that $c(l)/c(0) = 0.974$ ⁶². Using the economic data given in figure 5, the riskfree rate puzzle can be explained by $\lambda = 0.08$. In this case, $c^*/c_0 = 0.385$. Remember, for the equity premium to be explained for $\gamma = 3$, $d_l/d_h = 0.28$ (this is not a function of λ).

It is interesting to note that, because d_l/d_h is not a function of λ , concentrating the risk within a subset of the population does not help explain the equity premium. In particular, it will not be possible within this type of

⁶²Cecchetti et al. (1993) state that the chance of a "crash" is approximately 4 years in every 96, when dividends fall by 29.5% and consumption by 6.2%. We therefore appear to be underestimating the size of a crash here. Notice, though, that the size of the dividend crash is not enough to explain the equity premium for $\gamma = 3$.

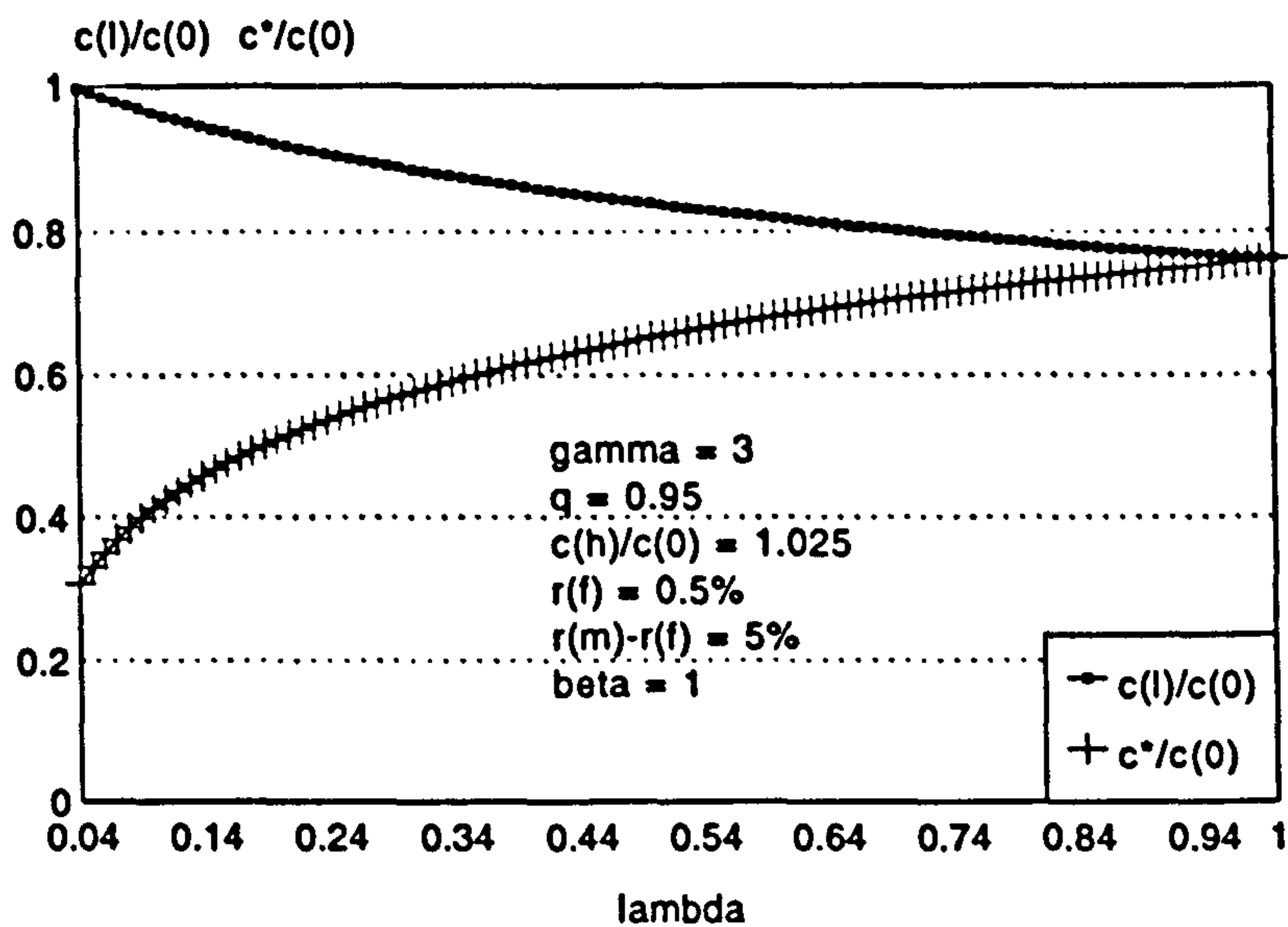


Figure 5: Two state economy, income risk concentrated in a subset of the population, size of risk needed to explain the Mehra & Prescott puzzles.

model to explain the equity premium using “realistic” data for $\gamma = 2$ since d_l/d_h is predicted to be negative for this coefficient of relative risk aversion (see figure 4).

13.3.2 Multiperiod models

In a one period model, unemployment is persistent. As the economy terminates at $t = 1$, it is not possible in this model to borrow or sell shares at times of unemployment with the intention of repaying the debt or buying back the shares when reemployed in future time periods. Clearly the amount that one would be prepared to borrow/sell at times of unemployment in multiperiod economies will depend on (i) the expected duration of the unemployment pe-

riod and (ii) the market frictions inhibiting trades in financial assets. These are the complexities that arise in multiperiod incomplete market models. The key work in this area has been undertaken by Deborah Lucas and John Heaton. Consider first an incomplete multiperiod economy where there are no frictions in trading financial assets. Lucas (1994) simulates two long-lived economy in which there is dividend risk and untradable labour income risk. In the first, the aggregate and individual risks are uncorrelated. In the second, the two sorts of risk are correlated in the manner of Mankiw (1986) — that is, unemployment can only occur in a low dividend state. However, in both economies the idiosyncratic element of risk is *iid*: that is the probability of being unemployed at $t + 1$ is independent of the employment status of the individual at t — there is no persistency of unemployment. In neither of the economies that she considers are the Mehra & Prescott puzzles resolved by the introduction of uninsurable financial risk. Over a lifetime positive endowment shocks will offset negative shocks to personal capital. Therefore if an investor has reduced income then all she need do is trade in financial markets (borrow or short sell equity) to raise funds which she can then pay off when she has an offsetting positive consumption shock. Therefore, in both these economies equilibrium asset prices are highly similar to the representative agent case. However, such policies allow “Ponzi games” — that is, there always a positive probability of the investor exceeding any finite borrowing level under this system (for a fuller discussion see, for example, Blanchard and Fischer (1989)). Due to the transitory nature of the individual risks,

however, Lucas finds borrowing and short-sales constraints rarely bind.

When the model permits unemployment to have persistence the possibility of an individual becoming bankrupt increase and introducing trading frictions into the model becomes more important. Heaton and Lucas (1992) and Heaton and Lucas (1993) consider economies with dividend risk, individual risk, transaction costs and borrowing/short sales constraints. These papers differ from each other in theoretical ways. The former is two-period⁶³, the latter long-lived. The former assumes idiosyncratic shocks to either be permanent or transitory, while the latter tries to accurately reflect the persistence of individual shocks estimated from US data. The former relates personal and aggregate shocks in the manner of Mankiw (1986) while the latter considers both this and the uncorrelated case. Despite these theoretical differences, though, the findings of the two models are very similar. First, given that individual shocks appear to be short lived in the real economy for the US, unemployment risk on its own is unlikely to explain the Mehra and Prescott puzzles. Adding in transaction costs to either the stock or bond markets, but not both, will also not explain away the puzzles. In this case investors will trade in the market with no transaction costs to smooth their individual income shocks. As (virtually) no trading need occur in the market with costs equilibrium prices are not affected by these costs⁶⁴. So, only by

⁶³They refer to it as three period as the economy has times $t = 0, 1, 2$. Strictly speaking, though, there are three points in time but only two time *periods*.

⁶⁴See Aiyagari and Gertler (1991) for related work. They consider an incomplete economy with no aggregate risk and exogenously determined returns to two riskless assets (one traded costlessly, one traded with cost) and predict asset/income ratios and transaction velocities. Again, the introduction of transaction costs into the economy does not help fully explain observed market behaviour.

adding transaction costs into both markets is their a potential explanation to the puzzle.

If costs are the same for buyers and sellers (borrowers and lenders) then both the supply and demand of financial assets is affected. The sellers require a higher price while the buyers require a lower price compared to the no transaction cost economy. These two counteracting effects offset and equilibrium asset prices return to approximately the complete market case. In stockmarkets, as buyers and sellers pay transaction costs, the introduction of such costs need not affect the predicted return to the market⁶⁵. However, in debt markets, it can be argued that only the borrower pays transaction costs. This asymmetry in transaction costs is crucial. The supply, but not demand, of debt instruments is effected reducing the riskfree rate (and hence increasing the equity premium). So, by having symmetric transaction costs in the stockmarket and assymetric transaction costs in the bond market the Mehra & Prescott puzzles might be resolved. Severe borrowing constraints are seen as being a form of assymetric transaction cost in the debt market and such constraints cause the most significant reductions in the riskfree rate. An excellent and more detailed review of this area (with a few minor original results) is given by Heaton and Lucas (1995) and the reader is referred to this paper for a more complete discussion.

What level of persistence is needed in income shocks to generate the

⁶⁵In fact the return to the market rises slightly as, with transaction costs, individual consumption will be more volatile than aggregate consumption which will increase the required excess return to the market while at the same time reducing the riskfree rate by increasing the precautionary demand for assets.

observed equity premium and real riskfree rate? Constantinides and Duffie (1996) try to find income processes that are consistent with observed asset prices and the Euler equation without trying to verify the plausibility of these income processes. It has been shown that the average equity premium and real riskfree rate can be resolved within a representative agent economy with power preferences, only the value of β, γ required for such an explanation are considered “unrealistic”⁶⁶. Let $\rho := -\ln\beta$. Consider the income at time t of agent k : y_{kt} . Restrict it to be of the following form:

$$y_{kt} = (y_t + d_t) \text{Exp} \left[\sum_{s=0}^t \left(y_s \eta_{ks} - \frac{y_s^2}{2} \right) \right] - d_t$$

Here y_t, d_t represent aggregate income and dividends and η_{ks} is a $N(0, 1)$ process specific to agent k that is independently distributed. y_s^2 measures the heterogeneity of the investment community. Introduce the new notation $s_t := \sum_{s=0}^t (y_s \eta_{ks} - \frac{y_s^2}{2})$. Notice that $E_{0,\eta}[e^{s_t}] = 1, E_{0,\eta}[e^{-\gamma s_t}] = e^{\sum_{s=0}^t ((\gamma + \gamma^2)/2) y_s^2}$.

Suppose that agents have power utility and an econometrician knows that. However, suppose that econometrician mistakenly believes that prices in the economy are as if a representative agent existed. Let the estimated parameters of the utility function under this assumption be $\hat{\gamma}, \hat{\rho}$ on the basis of aggregate data d_t, y_t . Let the real preference parameters be denoted by γ, ρ for each agent. If heterogeneity is restricted to be:

⁶⁶Although, as will be noted later in this chapter “In formal tests of the conditional Euler equations, Hansen and Singleton (1982) ... and others rejected the model even though no *a priori* upper bound is imposed on the relative risk aversion coefficient” (Constantinides and Duffie (1996) p. 220).

$$y_s^2 = \frac{2}{\gamma^2 + \gamma} [\rho - \hat{\rho} + (\gamma - \hat{\gamma})\Delta \ln c]$$

then the price, p_j of an asset that pays dividend d_{jt} at time t and zero otherwise is:

$$\begin{aligned} p_j &= e^{-\rho t} E_0[(d_{jt}(y_{kt} + d_t)/c_0)^{-\gamma}] \\ &= \frac{e^{-\rho t}}{c_0^{-\gamma}} E_{0,\eta} E_{0,d,t,d_t,y_t}[d_{jt}(d_t + y_t)^{-\gamma} e^{-\gamma s_t}] \\ &= \frac{e^{-\rho t}}{c_0^{-\gamma}} E_{0,d,t,d_t,y_t}[d_{jt}(d_t + y_t)^{-\gamma} e^{\sum_s ((\gamma + \gamma^2)/2) y_s^2}] \\ &= \frac{e^{-\hat{\rho} t}}{c_0^{-\gamma}} E[d_{jt}(d_t + y_t)^{-\gamma} e^{(\gamma - \hat{\gamma})(\ln c_t - \ln c_0)}] \\ &= \frac{e^{-\hat{\rho} t}}{c_0^{-\hat{\gamma}}} E_0[d_{jt}(d_t + y_t)^{-\hat{\gamma}}] \end{aligned}$$

In other words, the econometrician is unable to distinguish prices in a representative agent community with parameters $\hat{\rho}, \hat{\gamma}$ from one without a representative agent where individual preferences are parameterised by ρ, γ and an income process described by y_{kt}, y_s^2 . The degree of persistence in the income shocks required to reconcile the Mehra & Prescott puzzles to the Euler equation in a frictionless market has now been established. This work has been generalised by Saito (1993), who allows for idiosyncratic shocks and aggregate shocks to be correlated. Using realistic parameter estimates, he finds that the riskfree rate is well explained by this type of model. However, the predicted equity premium remains too low as the high precautionary savings motive in this economy increases the demand for all financial assets.

13.4 Savings behaviour of the unemployed

As emphasised in the previous section, in multiperiod models of incomplete markets it is the *persistence* of the idiosyncratic risk that is the crucial factor in determining equilibrium asset prices. In this thesis, particularly chapter 6, unemployment will be taken to be the key source of idiosyncratic risk. This subsection therefore considers the persistence of idiosyncratic risk in general and unemployment in particular. First we consider a representation of income where permanent and transitory components are smooth, then one where there are endowment shocks of moderate severity and finally unemployment.

For the smooth income process, the model of Carroll and Samwick (1996) is analysed. They assume that income is generated by the following process:

$$\begin{aligned}\ln(y_{it}) &= \ln(z_{it}) + \epsilon_{it} \\ \ln(z_{it}) &= \ln(z_{it-1}) + g_{it} + \eta_{it}\end{aligned}$$

Here y_{it} , z_{it} , g_{it} denote, respectively, total income to individual i at time t , the permanent component of that income and the predictable growth in income at $t - 1$. η_{it} , ϵ_{it} refer to the permanent and transitory component of income fluctuation. For professional and technical workers, total variance in annual income is 2.92% which can be decomposed into 1.72% permanent variation and 3.31% transitory variation. Standard errors are 0.39%, 0.62% and 1.16% respectively. Figures for the total population do not differ materially from this.

It is the work of Heaton & Lucas that has shown that the conclusions of Mankiw's single period model will not hold in a multiperiod environment.

Essentially they believe that endowment shocks are not sufficiently persistent. Using the PSID data, they divide investors equally into two categories. Each member of one category consumes 75% of average per-capita consumption while each member of the other group consumes 125% of average per-capita consumption. This captures the variance of the innovation of annual household income. They estimate the probability of switching category from one year to the next as 26% (see pp.6–7 of Heaton and Lucas (1995)). While this suggests that shocks are “relatively persistent” (*ibid.* p.7), the persistence is not enough to explain away the Mehra & Prescott puzzles unless there are severe market frictions.

Finally, we turn to severe idiosyncratic endowment crashes. In general, research in the US suggests that these crashes are short lived: “... income typically recovers from near-zero events within three years, and mostly recovers within a year” (Carroll (1992) footnote 19). However, it should be recognised that:

“...there is a huge variation in unemployment inflow rates and durations across countries. Unemployment durations are very low in North America, and inflow rates rather high. By contrast, in the EC inflow rates are quite low but durations are huge. And the ‘virtuous’ countries (Norway, Sweden, Finland and Japan) have both low inflow and low duration.”

Layard, Nickell and Jackman (1991) p. 222–4

Evidence presented in Layard et al. (1991) suggests that, of the UK unemployed, 35% are people with previous work experience who are experiencing a period of unemployed that has already lasted for more than three years (table 17, p.271). Of those unemployed for less than three years, the average

duration of unemployment is 12.8 months (table 3, p.45). Table 9 (p.422) shows that 1 in 4 people unemployed have been unemployed for more than 12 months (the LTU)⁶⁷. For the US, the percentage LTU is 4.2% (table 9, p.422), while the average duration of unemployment is 2.6 months (table 3, p.45). The duration of unemployment for professional and managerial staff is similar to the economy-wide duration in both countries (again table 3, p.45). These numbers may, though, overestimate the differences in duration of unemployment between the two countries as “one should, at any rate, be aware of the fact that roughly half of the unemployment spells in the USA end in withdrawal from the labour force rather than in a job” (*ibid.* p. 270). It would, though, seem reasonable to conclude that periods without work are more persistent in Britain than in the US. Therefore the one period model of Mankiw (1986) may be more applicable to the UK than the US.

Given that the potential threat of bankruptcy from following a strategy of borrowing/selling financial assets increases as the expected duration of unemployment increases, we would expect to see the British reluctant to trade assets when unemployed. The UK evidence is again presented by Layard et al. (1991):

“Do unemployed people run down financial assets or borrow to maintain their consumption, or do they simply consume less? ... In the 1978 British cohort study ... there was no evidence of savings being run down. Later British studies broadly confirm the 1978 results. Those aged under 35 in a sample of the 1983 inflow into unemployment did not, on average, reduce their savings or increase their borrowings

⁶⁷Notice that this is not consistent with the percentage given for those with more than three years continuous unemployment. This may be because people withdraw temporarily from the unemployment register without necessarily reentering work.

to limit the fall in their consumption during 15 months of unemployment, but those aged 35 and over increased their net debt by about £400 on average ... The 1987 cohort study shows that those unemployed for no more than nine months reported an increase in net debt from an average of £435 before becoming unemployed to £525 when they returned to work.”

ibid. p. 246–7

Gruber (1994) presents evidence on the savings of individuals prior to unemployment in the US: “There is some limited evidence on the savings behavior of the unemployed in the PSID...Among individuals who lose their jobs, only 56% had any savings before the job loss, and only 23% had savings of more than two months income. The comparable figures for those not losing their jobs were 84% and 52% respectively” (*ibid.* footnote 6). Therefore, unless the unemployed are prepared to build up large debts, there is a limit to how much consumption smoothing they can achieve through financial markets.

This is evidence on savings and not consumption, though. It may be that individuals are able to keep consumption at near employment levels even when unemployed and not run down financial assets through, for example, informal help from friends and family. Evidence on changes in log consumption on becoming unemployed in the US is given by Gruber (1994). Using PSID data from 1968–87, he estimates that average expenditure on food dropped by 6.8% on becoming unemployed, but the associated standard deviation of 42.4% suggests great cross-sectional variation. In the absence of unemployment benefit, Gruber predicts that the average drop in consumption would be around 22%. It should be noted from table 1 of Nelson (1994),

though, that only around $1/5$ th of total consumption is accounted for by food expenditure.

In conclusion, this evidence is a little confusing. It appears as if periods of unemployment are relatively short lived. The work of Heaton and Lucas would then imply that people should run down financial assets to maintain consumption during this period. It seems, though, that there is little, if any, decrease in saving (increase in borrowing) during periods of unemployment. At the same time, there also does not seem to be a large decrease in consumption on non-durables. This might be reconciled by supposing that individuals spend less on durable goods and / or get support from family and friends during periods of unemployment. Further empirical research is required to more accurately determine the persistence of risk to individual income.

14 Closely related puzzles

So far, this chapter has concentrated on developing, and then trying to explain, the Mehra & Prescott puzzles. As discussed in the previous chapter, the Euler equation is closely bound in with the CCAPM and, therefore, these puzzles should be interpreted in the light of broader CCAPM empirical tests. This literature is briefly reviewed in the next section. This section looks at two puzzles that are closely related to the Mehra & Prescott anomalies. First, an overview of Hansen-Jagannathan bounds tests is presented. This bound also comes directly from the Euler equation. Second we consider the

higher moments of asset returns. A valid asset pricing model should be able to explain the total behaviour of asset prices and not just the long term expectation.

14.1 Hansen-Jagannathan bounds

Suppose that there is a riskfree asset. From equation 3 we know that, for any asset i :

$$\begin{aligned} E[(r_i - r_f)\pi] &= 0 \\ \Rightarrow E[r_i - r_f]E[\pi] &= \text{Corr}(r_i, -\pi)\text{Std}(r_i)\text{Std}(\pi) \\ \Rightarrow \frac{E[r_i - r_f]}{\text{Std}(r_i)} &\leq \frac{\text{Std}[\pi]}{E[\pi]} \end{aligned}$$

This inequality must hold for all assets. Notice that the left hand side of this inequality is the Sharpe ratio. The asset with the greatest Sharpe ratio is the tangency portfolio on the mean-variance efficient frontier. By appealing to CAPM theory we can take the market portfolio to be the tangency portfolio and so the bound on $\text{Std}[\pi]/E[\pi]$ is given by $E[r_m - r_f]/\text{Std}(r_m)$. On the Mehra & Prescott data the Sharpe ratio of the market portfolio is 0.374. Taking annual consumption data from Grossman et al. (1987) π to be the ratio of marginal utilities and the usual power utility function with $\beta = 0.97$, this bound is violated with $\gamma \leq 6.25$.

The alternative is to assume that there is no riskless asset. The real rate of return on treasury bills has a measurable standard deviation. So short term government securities can be included in a variance / covariance matrix of asset returns. In this case, take the vector version of the Euler equation, $p(\mathbf{x}) = E[\pi\mathbf{x}]$, for a vector of payouts \mathbf{x} . The bound of Hansen

and Jagannathan (1991) follows from this. The discussion here is based on that given in Cecchetti, Lam and Mark (1994). Regress $\pi - E[\pi]$ onto the horizontal vector $(\mathbf{x} - E[\mathbf{x}])^T$ whose i^{th} element is $x_i - E[x_i]$. The regression vector will be denoted by \mathbf{z} and the error term, which is orthogonal to $\mathbf{x} - E[\mathbf{x}]$, by u . $\Sigma_{\mathbf{x}} = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T]$ is the variance, co-variance matrix of payoffs (not returns).

$$\begin{aligned}
\pi - E[\pi] &= (\mathbf{x} - E[\mathbf{x}])^T \mathbf{z} + u \\
\Rightarrow E[(\mathbf{x} - E[\mathbf{x}])(\pi - E[\pi])] &= E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T] \mathbf{z} + E[(\mathbf{x} - E[\mathbf{x}])u] \\
\Rightarrow \mathbf{z} &= \Sigma_{\mathbf{x}}^{-1} E[(\pi - E[\pi])(\mathbf{x} - E[\mathbf{x}])^T] \\
\Rightarrow \mathbf{z} &= \Sigma_{\mathbf{x}}^{-1} (E[\pi \mathbf{x}] - E[\pi]E[\mathbf{x}]) \\
\Rightarrow \mathbf{z} &= \Sigma_{\mathbf{x}}^{-1} (p(\mathbf{x}) - E[\pi]E[\mathbf{x}])
\end{aligned}$$

Now, consider the variability of π . $\sigma_{\pi}^2 = E[(\pi - E[\pi])^2] = E[((\mathbf{x} - E[\mathbf{x}])^T \mathbf{z} + u)^2] = E[(\mathbf{x} - E[\mathbf{x}])^T \mathbf{z}]^2 + E[u^2]$ as u is orthogonal to $\mathbf{x} - E[\mathbf{x}]$.

Now, $E[u^2] \geq 0$, and so

$$\begin{aligned}
\sigma_{\pi}^2 &\geq E[(\mathbf{x} - E[\mathbf{x}])^T \mathbf{z}]^2 \\
&= E[\mathbf{z}^T (\mathbf{x} - E[\mathbf{x}]) (\mathbf{x} - E[\mathbf{x}])^T \mathbf{z}] \\
&= \mathbf{z}^T E[(\mathbf{x} - E[\mathbf{x}]) (\mathbf{x} - E[\mathbf{x}])^T] \mathbf{z} \\
&= (p(\mathbf{x}) - E[\pi]E[\mathbf{x}])^T \Sigma_{\mathbf{x}}^{-1} \Sigma_{\mathbf{x}} \Sigma_{\mathbf{x}}^{-1} (p(\mathbf{x}) - E[\pi]E[\mathbf{x}]) \\
\Rightarrow \sigma_{\pi} &\geq [(p(\mathbf{x}) - E[\pi]E[\mathbf{x}])^T \Sigma_{\mathbf{x}}^{-1} (p(\mathbf{x}) - E[\pi]E[\mathbf{x}])]^{0.5} \\
&= [(1 - E[\pi]E[(1 + \mathbf{r})])^T \Sigma_{\mathbf{r}}^{-1} (1 - E[\pi]E[(1 + \mathbf{r})])]^{0.5}
\end{aligned}$$

Where $\mathbf{1}, \mathbf{r}$, as usual, denote the vector of 1s and of returns to the risky asset and the covariance matrix now refers to returns rather than payoffs. Recall the discussion above on the relationship between the Hansen-Jagannathan bounds and the mathematics of the tangency portfolio. Divide the previous offset equation left and right by $E[\pi]$. Remember that the current environment is no riskfree asset, but, if there were, $1/E[\pi] = 1 + r_f$.

The righthand side is equivalent to $E[r_T - r_f]/\text{Std}(r_T)$ where r_T refers to the return on the tangency portfolio. This is easily shown using equations (19), (20), p.89 of Ingersoll (1987). So, the Hansen-Jagannathan bound is closely related to mean-variance efficiency. In the first representation (with a riskless asset) it is assumed that the market index is mean variant efficient. In the second representation (no riskless asset), the tangency portfolio is explicitly calculated using $E[\pi]$ and the variance/covariance matrix.

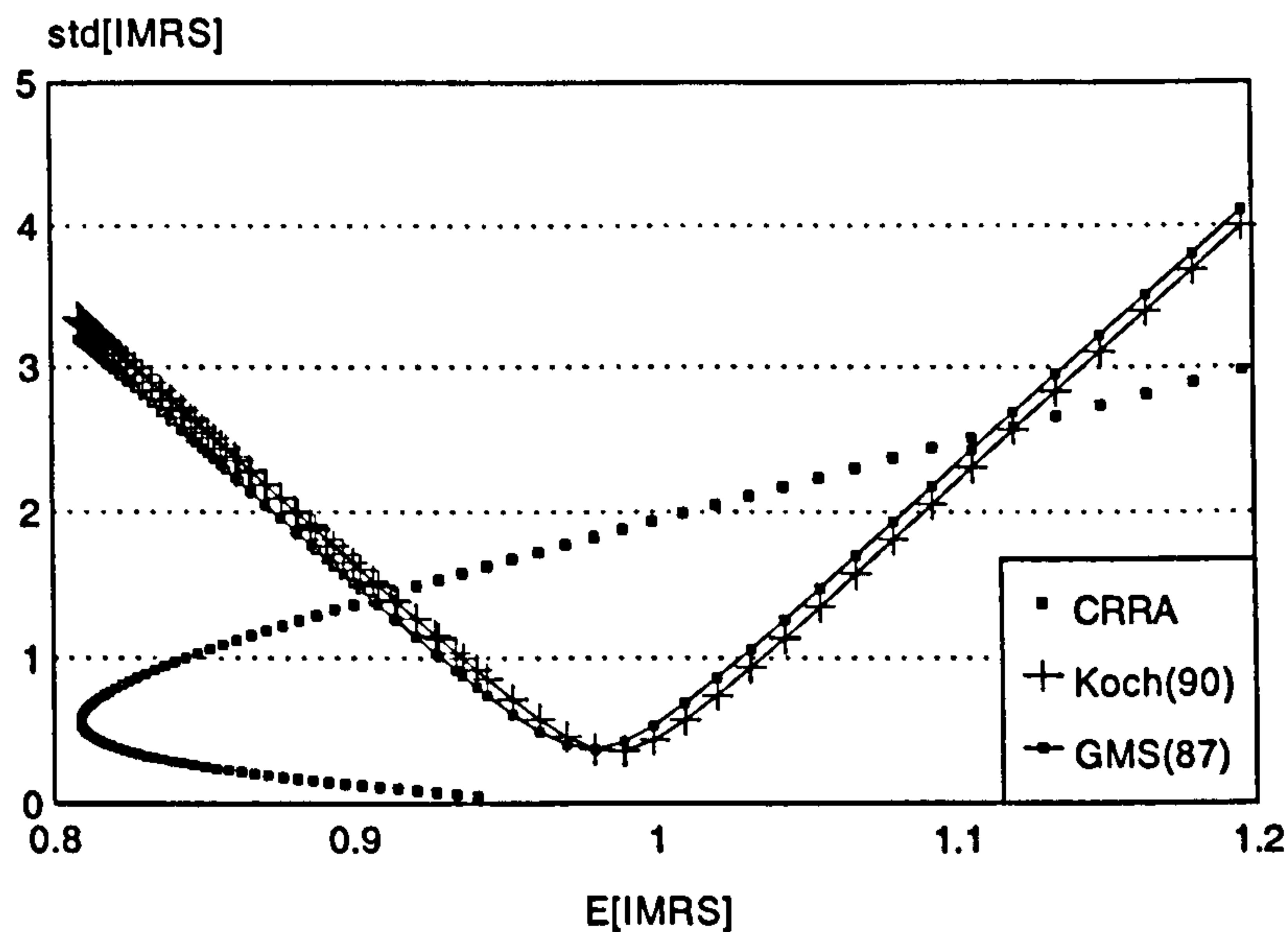


Figure 6: The Hansen-Jagannathan bound tested on annual consumption data from the US 1890-1980: source Grossman, Melino & Shiller (1987). The “CRRA” points give the mean and standard deviation of the IMRS using this consumption data with a CRRA utility function for $\beta = 0.97, \gamma \in [1, 25]$. As γ rises, so the standard deviation rises. The two solid lines give the Hansen-Jagannathan bound for this mean / standard deviation pair. The Koch(90) line uses the average real returns and (co)variances to the market and a riskfree asset for the US over the interval 1888–1978 as given by Kocherlakota (1990b). The GMS(87) line uses the real returns and (co)variances to the market, a short term riskfree asset and a long term bond index over the interval 1890–1980 as given by Grossman, Melino & Shiller (1987).

This restriction can now be tested on observed data. Initial tests concentrated on two assets: the market index and a treasury bill. A graph of the Hansen-Jagannathan bound is produced in figure 6. Consumption data was taken from Grossman et al. (1987) for the interval 1890–1980. It was assumed that utility takes the usual time-separable, constant relative risk aversion form with coefficient of impatience $\beta = 0.97$. The coefficient of relative risk aversion was allowed to change from 1 to 25 (step size 0.25). Two bounds were created. One uses the annual 1888 – 1978 Mehra and Prescott (1985) data as presented by Kocherlakota (1990b), who quotes⁶⁸ $E[r_f, r_m] = [0.010, 0.070]$, $\sigma[r_f, r_m] = [0.055, 0.165]$ and $\text{Corr}[r_f, r_m] = 0.114$. This bound is, then, identical to that given in figure 1 of Burnside (1994) (although over a slightly wider domain), but differs slightly from figure 1 of Hansen and Jagannathan (1991), who use raw data from Campbell and Shiller (1988). The second bound is created using three series of asset returns (stocks, long bonds and treasury bills) from 1890–1980 as given by Grossman et al. (1987). As can be seen, the two bounds are very similar. Only when γ exceeds 17 are both bounds satisfied.

Notice that, while violations of the Hansen-Jagannathan bound is “similar” to the Mehra & Prescott puzzle, it differs in an important way. In the development of the equity premium puzzle one of the key independent variables was the covariance between the return to the market and the pricing kernel. To draw the Hansen-Jagannathan bound the variance/covariance structure

⁶⁸This notation is a little confusing. “ r_f ” is used to denote the return to treasury bills, which is no longer assumed to have zero variance and so is not, strictly speaking, riskfree.

of asset returns is required but the covariance between asset returns and the pricing kernel is not. Therefore models that try to explain the equity premium puzzle by increasing the correlation of market and consumption risks will not, on their own, help explain any violation of the Hansen-Jagannathan bound.

The reason why, for power utility, the mean / standard deviation pairs of IMRS are curved is not initially clear. In the single period, non parametric development of the equity premium puzzle it was shown that $\pi \approx \beta(1 - \gamma \Delta \ln c)$. Taking expectations of both sides, it is clear that, under this approximation $E[\pi]$ should be monotonic decreasing in γ . As figure 6 shows, though, above certain values of γ $E[\pi]$ increases. This is because of important second order terms, captured in the bottom two equations for the real riskfree rate given in equation 35 that reflect the precautionary savings motive (see above).

Figure 6 obviously provides spot estimates as to whether the mean/standard deviation pairs for a certain utility function parameterised on certain data fits the IMRS bound calculated from financial returns. It provides no evidence as to the statistical significance of the violation of the bound. Such examination is a detailed econometric issue and is considered to be outside the scope of this thesis. The reader is referred to, for example, Ferson (1995), Burnside (1994) for a review of this issue. There are also various amendments that can be made to the basic Hansen-Jagannathan bound. First, the basic test places no restrictions on the strict positivity of π . This can be done using numerical procedures that provide a sharper bound. Snow (1991) ar-

gues that this bound also loses important data concerning higher moments of asset returns. Using the restriction that $1/\delta + 1/q = 1$, Snow exploits the relationship $1 = E[\pi(1 + r)] \leq E[\pi^\delta]^{1/\delta} E[(1 + r)^q]^{1/q}$ (Hölder's inequality). Therefore there is a general restriction on the δ^{th} moment of the IMRS in terms of the q^{th} moment of asset returns. The case $\delta = q = 2$, the Hansen-Jagannathan bound, is a specific case of this general rule. Snow examines the general bound for $\delta = 2, 3/2$ and 3 using numerical techniques. See also He and Modest (1992) for an analysis of how transaction costs and short selling / borrowing constraints affect the bounds.

The aim of this section is not to provide a comprehensive review of empirical tests of Hansen-Jagannathan bounds. Instead, the aim is to alert the reader to the fact that any “true” potential explanation to the Mehra & Prescott puzzles must also pass the Hansen-Jagannathan bounds tests. For example, Hansen and Jagannathan (1991) themselves show that habit formation models fit their bounds more closely than time-separable models. Of particular relevance here is Telmer (1993). Telmer works in the environment of Mankiw (1986), which is central to the models in later chapters (particularly chapter 6) incorporating borrowing constraints. There is one financial asset, a riskless bond, through which agents can partially smooth their endowment fluctuations. As endowment shocks are transitory in his model, adding idiosyncratic endowment risk does not help explain violations of the Hansen-Jagannathan bound unless borrowing restrictions are very severe. This, therefore, is highly similar to the results of Heaton & Lucas

reviewed above for the equity premium puzzle. Of course, if endowment risk is persistent in the way described by Constantinides and Duffie (1996) then any bound can be satisfied.

14.2 (Co)Variances of asset returns

If adjustments to standard consumption based pricing models to account for the puzzles of Mehra & Prescott are robust then, not only should they satisfy the Hansen-Jagannathan bound, but they should also explain higher moments of asset returns. In particular, the variance/covariance matrix of asset returns should be also be explained by a valid asset pricing model⁶⁹. This is a more stringent test than Hansen-Jagannathan bounds (Cecchetti et al. (1993)). In general the adjustments to the standard model presented in this chapter cannot explain both the first and second moments of the excess return to the market and real rate of return on treasury bills. See Heaton (1995) for evidence on habit persistence (and his own mixed persistence / durability model), Kandel and Stambaugh (1991) for recursive utility functions, Lucas (1994) for incomplete market models and Cecchetti et al. (1993) for Markov models.

The problem is as follows. In order to have a high equity premium and a high volatility in observed excess returns, the intertemporal marginal rate of substitution needs to be volatile. However, as soon as a highly volatile IMRS is included in the model the volatility of the riskfree rate becomes too

⁶⁹Other characteristics of asset markets could also be included — particularly autocorrelation in asset returns and transaction velocities.

great. That is, the difficulty arises because the real riskfree rate is so *smooth* while the equity premium is so high and stock returns are so volatile. To quote Heaton (1995): “... the requirement that there is little volatility in the bond returns constrains the model from fitting the equity premium and the volatile stock returns” (*ibid.* p.706 — see also the concluding sentence in Cecchetti et al. (1993)). So not only must the low average riskfree rate, high average equity premium be explained, but also the low volatility of the riskfree rate and high volatility of equity returns must also be explained.

15 Empirical tests of consumption based asset pricing models

In this chapter so far, the Mehra & Prescott puzzles have been developed and various potential explanations have been presented. This section does not aim to provide detailed econometric analysis, nor indeed be comprehensive in its treatment, of the numerous tests of the CCAPM that have been conducted as this dissertation has, in general, a more theoretical than statistical approach. The reader is referred to, for example, excellent reviews of this area that are given by Singleton (1990) and Ferson (1995) for more comprehensive coverage. Instead this section aims to address two questions by examining the main papers in the area. First, can the cross-sectional variation in asset returns be reconciled with the complete market version of the Euler equation/CCAPM under power utility and, if so, what parameter values for β, γ might be reasonable.

An early estimate for the coefficient of relative risk aversion of a representative agent was provided by Friend and Blume (1975) who conclude that it is almost certainly more than 1 and probably more than 2. Grossman and Shiller (1981) essentially test an Euler equation version of the variance bound of Shiller (1981) and LeRoy and Porter (1981). Grossman and Shiller used the multiperiod Euler equation, with “perfect foresight” of consumption and dividends to try to simulate the behaviour of the market index. They find that for $\gamma = 4$ they can find reasonable correlation between predicted asset returns and the observed values. In tests of the arbitrage pricing theory Chen, Roll and Ross (1986) find that consumption is not statistically significant as an explanatory factor for the US (Poon and Taylor (1991) do not include consumption as a variable in their UK based test of the APT). Aside from these, there are essentially four approaches to testing these consumption based models. The first is to assume some parametric form for consumption and asset returns and test the models restrictions under these parameterisations. The first paper to take such an approach was Hansen and Singleton (1983). The second method is to make no assumptions about the distribution of consumption and asset returns and use a non-parametric technique. This was initially used in this context by Hansen and Singleton (1982). Brown and Gibbons (1985) try to adjudicate on which of these two approaches is more satisfactory. They find (using a simplification of the Euler equation that requires only market index and riskfree rate data), that there is very little difference in the precision of their parameter estimates between the

parametric and non-parametric tests. Given the increased robustness of the latter, they conclude that more emphasis should be given to non-parametric than parametric tests⁷⁰. A third approach initially advocated by Mankiw and Shapiro (1986) is to test the relationship between realised returns and consumption betas. Some authors (notably Breeden, Gibbons and Litzenberger (1989)) argue, though, that this underestimates the predictive power of the model as consumption data is observed discretely and yet the model is continuous. The fourth approach, therefore, is to adjust spot consumption data to account for the fact that it is non-continuous and test the model under the adjusted consumption data.

15.1 Parametric tests

In the previous chapter it was shown that, under a single period parametric representation of asset returns and consumption (equations 29, 26):

$$E[R_i] + 0.5\text{Var}(R_i) = -\ln(\beta) + \gamma(E[\Delta \ln c] + \text{Cov}[R_i, \Delta \ln c] - 0.5\gamma^2 \text{Var}(\Delta \ln c)) \quad (40)$$

$$E_t[R_{jt} - R_{it}] = \frac{\text{Var}(R_{it}) - \text{Var}(R_{jt})}{2} + \gamma \text{Cov}(R_{jt} - R_{it}, \Delta \ln c_t) \quad (41)$$

This equation forms the basis for analysis of Hansen and Singleton (1983), Ferson (1983), Hall (1988) and Wheatley (1988). Notice, though, that it is not possible to observe $E[\Delta \ln c]$ directly. Three ways have been adopted to overcome this problem. First, one can take survey data on expected changes

⁷⁰It is worth mentioning that Brown and Gibbons (1985) estimate the coefficient of relative risk aversion to be between 1 and 2, which shows reasonable intertemporal stability.

in variables as a proxy for average investor expectations. Hall (1988) uses this approach for part of his paper. The second technique is to estimate the expectation using instrumental variables that are available to the econometrician at time $t - 1$. Following Hansen and Singleton (1983), all four papers in this area assume that $E[\Delta \ln c]$ can be expressed as a polynomial in lags of log asset returns and log consumption changes. Notice that, from the previous equation, expected changes in log asset returns are proportional to expected changes in log consumption changes, with the coefficient of proportionality being the coefficient of relative risk aversion. Therefore, the assumption that expected changes in log consumption can be expressed as a polynomial in lagged operators of log asset returns and log consumption changes imposes a similar property on expected changes in log returns. Using maximum-likelihood methods, these papers are then able to jointly estimate these lag polynomial parameters as well as γ, β (under assumptions of stationarity). They can also test the accuracy of the null hypothesis of the equation 40 by using a likelihood ratio test statistic to look for overidentification of the model⁷¹. The third approach is to deal with equation 41 direct by observing that $R_{it} - R_{jt}$ is orthogonal to all variables in the econometrician's observable data set at $t - 1$. Therefore $R_{it} - R_{jt}$ can be regressed against lagged asset returns with the null hypothesis being that the regression coefficients should be zero. This approach, though, does not

⁷¹A discussion of the full econometric issues are outside the scope of this paper. The reader is referred to, for example, Davidson and Mackinnon (1993) for a discussion on instrumental variables, overidentification (both chapter 7) and likelihood ratio test statistics (chapter 8).

provide estimates for γ, β .

Using monthly data from 1959 to 1978 and a value weighted market index Hansen and Singleton (1983) find γ to be between 1 and 2 and estimate β to be below, but close to one. Under this specification, they were unable to reject the general model. However, under alternate specifications of the model (using individual share data, riskfree returns or the difference between asset returns) the model appeared to perform badly. Wheatley (1988) confirms that treasury bill returns and consumption data do not appear to corroborate these parametric specifications of the Euler equation. Through the use of simulations, though, he argues that maximum likelihood estimators of high coefficients of relative risk aversion are too low in small samples. This is very closely related to the work of Kocherlakota (1990b), that finds coefficients of relative risk aversion can also be underestimated using non-parametric approaches. The main contribution of Wheatley (1988) is, that he contends that if consumption data is measured with error then the true relationship between consumption and asset returns will be underestimated by applying econometric techniques to reported data. This leads to overrejection of consumption based asset pricing models. Again using simulations he shows that the model can be rejected 50% of the time at the 1% level using “reasonable” estimates of measurement error. Ferson (1983) observed that equation 40 holds not only under the assumption of power utility, lognormal asset returns and consumption changes but also for constant absolute risk averse utility, changes in normal real consumption and lognormal as-

set returns. Therefore, this equation can be used to test constant relative risk aversion against constant absolute risk aversion utility functions. Using quarterly data from 1947–1980 he finds very little evidence of predictable patterns in changes in consumption using inflation, stock market returns and consumption data as instrumental variables. However, there are predictable patterns in changes in the real returns on 3-month treasury bills over the period. This is difficult to reconcile with the observation that expected asset returns and expected log consumption changes should have the same characteristics in terms of the lagged polynomial on instrumental variables under the null hypothesis. Unsurprisingly, therefore, Ferson (1983) has difficulty in reconciling his consumption/t-bill data with the Euler equation. Two out of three estimates of relative risk aversion are negative in the case of CRRA. He suggests this might be due to time variation in the covariance terms — that is, a violation of the assumptions of stationarity.

15.2 Non-parametric tests

The non-parametric approach relies heavily on the Generalised Method of Moments (GMM) approach as developed by Lars Peter Hansen (1982). Again, a full review of this technique is outside the scope of this thesis and the reader is referred to one of many text book treatments of this area⁷². A brief discussion of the intuition behind the tests follow. It is known from the Euler equation that for all assets i , $E_0[(1 + r_i)U'[c_1, 1]/U'[c_0, 0] - 1] = 0$. There-

⁷²Again Davidson and Mackinnon (1993) provides a good review, while Hamilton (1994) explicitly refers to Hansen and Singleton (1982) in its discussion of GMM. Ferson (1995) reviews this area from more of a financial economics perspective.

fore, $(1 + r_i)U'[c_1, 1]/U'[c_0, 0] - 1$ is orthogonal to all data observable to the econometrician at time 0. Therefore, it is possible for the econometrician to introduce a set of instrumental variables available at time $t = 0$ and test for orthogonality between these variables and $(1 + r_i)U'[c_1, 1]/U'[c_0, 0] - 1$. If (s)he is going to observe the returns to x assets and there are z instrumental variables then there are xz orthogonality conditions that can be tested. The econometrician is trying to estimate two parameters (in the case of isoelastic utility), β, γ , and GMM provides a technique to identify the values of these parameters that comes “closest” to fulfilling all xz orthogonality conditions. However, as most tests have many more than two orthogonality conditions, it is possible for the model to be overidentified. Therefore GMM also provides an overidentification diagnostic to test whether the model itself can be rejected by the model.

There are two main papers in this area. Hansen and Singleton (1982) (and errata — Hansen and Singleton (1984)) test isoelastic utility under GMM using monthly data from February 1959 to December 1978. They use three series of asset returns and various series of instrumental variables to implement GMM. They estimate γ to be somewhat less than one and β to be below, but very close to one. Their tests of overidentification depend on the instrumental variables used in the test. However, for certain sets of instrumental variables the model is clearly overidentified while GMM should not be overidentified for any viable instrumental variable sets. So, while the parameter estimates are reasonable, the support for the model is not strong.

Epstein and Zin (1991) use GMM to test the Kreps-Porteus utility function. Using monthly data from April 1959 to December 1986 they estimate that the elasticity of substitution is always less than one, the coefficient of relative risk aversion is about one and beta is greater than one. As with Hansen and Singleton (1982) they find that the choice of instrumental variables significantly influence their results. The overidentifying restrictions again suggest that the model is often overidentified. Ferson and Merrick (1987) uses, as its instrumental variable, a dummy variable indicating whether the economy is in a recessionary or non-recessionary period. The coefficient on this variable is significantly different from zero which can only be reconciled with the Euler equation if parameters β, γ are different in recession and non-recessionary periods. So, these non-parametric tests cannot be considered to give strong support for standard models⁷³.

“The (GMM) test is based on a χ^2 statistic that summarises, in one number, how the data conform to the model’s many restrictions. The tests usually reject. This is not surprising since we know all models are false. The disappointment comes when the rejection is not pursued for additional descriptive information, obscure in the χ^2 test, about which restrictions of the model (time-series, cross-sectional or both) are the problem. In short, tests of the consumption model sometimes fail the test of usefulness; they don’t enhance our ability to describe the behaviour of returns.”

Fama (1991) p. 1596

Kocherlakota (1990a), in the vein of Wheatley (1988), conducts a simulation using GMM to test the accuracy of the parameter estimates and

⁷³See also Chan, Foresi and Lang (1996) who produce empirical evidence (including GMM tests) which gives greater support to their money-based CAPM than the CCAPM. The tests are not really supportive of either model and in both cases the estimates of relative risk aversion are both very high and imprecise.

overidentifying restrictions⁷⁴. In order to match the mean return and variance of the riskfree rate and equity premium over the period 1889 – 1978 she sets $\beta = 1.139, \gamma = 13.7$ (notice that these values of β, γ that solve the puzzle). Using GMM and four hundred simulations of ninety data points she finds that the point estimates of β, γ are reasonably close to their true values. However, the overidentifying restrictions reject the model much too frequently. She argues that the Hansen and Jagannathan (1991) bounds test is less prone to overrejection.

15.3 Realised returns vs. consumption betas

Mankiw and Shapiro (1986), Wheatley (1988) (for the US) and Sauer and Murphy (1992) (for Germany) estimate the CCAPM using an approach to see whether there is a positive relationship between consumption betas and realised returns for assets. They also compare the predictive abilities of the CAPM against those of the consumption CAPM. As mentioned in chapter 2, under Kreps-Porteus utility, predicted asset returns have both a CCAPM and CAPM component. Using quarterly data from 1959–82, Mankiw and Shapiro (1986) find “no support” for the CCAPM in isolation and discover that, when comparing the CAPM and CCAPM, “the coefficient on the market beta is always far larger and far more significant than is the coefficient on the consumption beta” (*ibid.* p. 457). Sauer and Murphy (1992) support these findings for quarterly data (and monthly data using the MCP method of

⁷⁴She also shows that the technique of Friend and Blume (1975) provides very misleading (understated) estimates of risk aversion.

Breeden et al. (1989) outlined below) from 1968–88 in Germany. Returning to the US, Wheatley (1988) finds, using monthly data from 1959–81, a positive, and significant, coefficient on the beta coefficient. However, the estimate of relative risk aversion is very high (over 100). Therefore, while the model cannot be rejected, the estimates of risk aversion appear to be unrealistic.

15.4 Adjusting spot consumption data

It can be argued that tests that use quarterly data to test the CCAPM will understate the predictive power of the model. This is because quarterly consumption is less volatile than spot consumption. It can be shown (see, for example, Breeden et al. (1989)) that the variance of quarterly consumption is $2/3$ the variance of spot consumption and consumption betas will be estimated that are $3/4$ their true “spot” values. Moving to monthly data significantly reduces this problem — the estimated consumption beta now becomes 93.75% its true spot value. Unfortunately monthly consumption data is not available over the long term and therefore the major empirical studies of the CCAPM have used quarterly data for their analysis. Grossman et al. (1987) and Breeden et al. (1989) address this issue in different ways⁷⁵. The latter does not attempt to measure betas against consumption *per se*, but against a portfolio of stocks that has maximum correlation with quarterly consumption data (the “maximum correlated portfolio” (MCP)). Such an approach is theoretically justified by Breeden (1979) and Breeden et al. (1989) themselves. The consumption CAPM can then be tested using

⁷⁵See also Litzenberger and Ronn (1986) for an early paper in this area.

monthly stock market data. Using data from 1929–82 and implementing a maximum likelihood estimation approach, they find that higher consumption betas earn higher returns and the relationship between risk and return appears reasonably linear. This paper can be considered to give qualified, but by no means overwhelming, support for the general consumption based asset pricing model. Grossman et al. (1987) postulate a joint continuous time form for consumption and asset returns that predict that samples of these variables taken at regular intervals should describe an ARMA(1,1) process. Using a number of different data sets over the period 1890–1981 and using a maximum likelihood estimation approach they use point values of asset and consumption data to estimate the parameters of the continuous time process. Using this process, they find that the model is both overidentified and get a wide range of estimates of relative risk aversion from around 2 to well over 100.

16 Conclusion

This chapter has demonstrated that the representative agent consumption based asset pricing theory developed in chapter 2 is unable to explain the average equity premium and real riskfree rate in the US over the last century. The theory “fails” because aggregate consumption has grown so fast and so smoothly. As the Mehra & Prescott puzzles can be presented under a number of different specifications of the representative agent paradigm, this restricts the number of potential solutions to the puzzles: four paths

to reconciling the theory and evidence have been presented here. First, the *ex-post* realisation of aggregate consumption and asset returns may not be representative of *ex-ante* expectations. The crash model of Rietz (1988) and survivorship evidence of Brown et al. (1995) might suggest that investors were worried in advance about potential market / consumption crashes that have not been realised in this century in the United States. This would violate the assumption of smooth aggregate consumption that was needed to derive equation 35. The low precision of estimates of average asset returns should also be emphasised: an average equity premium of 4% and real riskfree rate of 2.5% lies within the 95% confidence bound of Cecchetti et al. (1993). Second, by relaxing the assumption of additively time-separable power utility functions, the coefficient of relative risk aversion can be separated from the elasticity of intertemporal substitution. While allowing for habit persistence and recursive utility functions helps explain the riskfree rate puzzle the equity premium puzzle is surprisingly robust to more general specifications of investor preferences. Third, it has been suggested that incomplete markets may solve the puzzles. The impact of market incompleteness is amplified by a fourth relaxation of the representative agent paradigm — market frictions and borrowing constraints. It has been argued, both in a precautionary savings context and in models with both risky and riskfree assets, that the key driver of such models is the *persistence* of idiosyncratic risk. If individual risk is long lived, then it is optimal for investors to prepare themselves in advance for such risk. This has significant impact on equilibrium asset

prices. If endowment risks are short term then financial markets can be used to smooth consumption. Only in the presence of severe market frictions in all financial markets will asset prices deviate significantly (in the economic sense) from the complete market case. The incomplete market models that appear most likely to resolve the Mehra & Prescott puzzles are ones where there is a low probability, high impact, endowment “crash” (unemployment) that is long lived. Evidence for the US, though, suggests that unemployment shocks are reasonably short term, although periods of unemployment in the UK are longer lived. This has lead several authors (see, for example, Heaton & Lucas and Aiyagari (1994)) to conclude that incomplete market models are unlikely to explain the puzzles of Mehra & Prescott unless there are fairly strong market frictions in both the stock and bond markets. This evidence is not, though, conclusive and in chapter 6 an innovative empirical test is run that aims to determine the importance of idiosyncratic endowment shocks to equilibrium asset prices.

Following this lengthy debate on the Mehra & Prescott puzzles, the chapter finishes with two sections on related empirical tests. Section 14 considers Hansen-Jagannathan bounds tests and the ability of various models to explain higher moments of observed asset returns. As no model has been presented that is fully successful in explaining the first moments, it is unsurprising that the models presented earlier in the chapter do not pass these more stringent tests. Section 15 briefly highlights the fact that the Mehra & Prescott puzzles are by no means the only empirical anomalies that result

from standard representative agent consumption based asset pricing theory. Several studies have been able to reject the model while none have found strong support for it. Estimates of the coefficient of risk aversion have varied widely and in some cases have even been negative. The standard errors of the point estimates quoted are also often very large. Because of this, some authors are now looking to associate the pricing kernel with other macro variables than aggregate consumption. For example, Cochrane (1991a) believes that production variables influence equilibrium asset pricing while Chan et al. (1996) argues that the growth in monetary aggregates (such as M2 or M3) is more likely to explain asset returns than aggregate consumption growth.

This complete the main literature review section of the thesis. The next three chapters proceed as follows. In this chapter, the one period incomplete market model of Mankiw (1986) was presented. In this model, personal endowment risks are high when dividends are low. The next chapter looks in detail at the model of Weil (1992a), where dividend risks and endowment risks are independent. Two major contributions are made. First, an integrated approach is taken to the highly relevant theoretical literature on proper risk aversion. This provides a framework for linking several papers in the area. A new type of proper risk aversion, called “Basic Risk Aversion”, emerges naturally. The second major contribution is that this chapter highlights the extreme sensitivity of estimates of the effect of unemployment on asset returns to the exact specification of the utility function used. Chapter 4 makes it clear that further work is urgently needed on estimating the

behaviour of investor preferences in low endowment states. In chapter 5, optimal aggregate dividend policy is considered in an economy with idiosyncratic endowment shocks. It is argued that investors are averse to dividend cuts (at the aggregate level) at times of high income risk. This may explain the observed tendency for dividends to be smoothed and for rights issues to be concentrated in bull markets. This is, to the author's knowledge, the first application of the incomplete market theory discussed in this chapter to corporate finance issues. Chapter 6 looks at the impact of unemployment shocks on stock and treasury bill returns. If unemployment is the key factor determining the long term average equity premium and real riskfree rate then we might expect unemployment news to be a key state variable determining changes in asset prices. That is, chapter 6 makes a contribution to asset pricing theory by examining how *changes* in asset prices are affected by *changes* in expectation of unemployment. It is concluded that it is difficult to reconcile the rise in the riskfree rate prior to "bad" unemployment news with the precautionary savings motive.

Part IV

Proper risk aversion and the Mehra & Prescott puzzles

Proper risk aversion and the Mehra & Prescott puzzles⁷⁶

Abstract

This chapter examines Weil's application of proper risk aversion to the puzzles of Mehra & Prescott and particularly the equity premium anomaly. The qualitative effects of proper risk aversion are captured by decreasing absolute risk aversion and decreasing prudence. This chapter examines the quantitative effect of including independent exogenous income risks in equilibrium asset pricing models. If it is assumed that the utility function can be described by an equation with fixed parameters across its domain then it is shown that the model's predictions may vary considerably depending on the properties of very high derivatives (fifth and above) of the equation. This chapter also provides an integrated approach to local proper risk aversion.

⁷⁶Versions of this chapter were presented to the Accounting and Finance group, University of Warwick, February 1996 and as part of the Doctoral Finance Seminar Series, University of Warwick, May 1996. I am grateful to the participants for their useful comments. I would also like to thank Philippe Weil and Christian Gollier for their helpful observations on this chapter.

17 Introduction

In the last chapter, the puzzles of Mehra & Prescott were described in detail and potential solutions to these anomalies were outlined. The role that uninsurable income risk might play by increasing the volatility of individual consumption is highlighted. In particular it is contested that the one period model of Mankiw (1986), where the cross-sectional variance in individual income is negatively correlated with the dividend paid by the market, is capable of resolving the puzzles⁷⁷. In this chapter, a one period model is also applied to the Mehra & Prescott puzzles but in this case income risk is *independent* of marketable risk. Such an approach was initially taken by Weil (1992a) and his paper forms the basis for the current study.

The intuition is as follows. Mehra and Prescott realised that the observed average equity premium can be explained if the coefficient of relative risk aversion of all investors is much higher than is generally accepted. Pratt and Zeckhauser (1987) have recently started a debate which considers the role that independent exogenous risk to personal capital plays in the determination of risk aversion. An investor whose aversion to a marketable risk increases on the introduction of a background external income risk is called “proper risk averse”. Recent papers, notably Kimball (1993), show the change in an investor’s aversion to a marketable risk on the introduction of background uncertainty is driven by the second to fourth derivatives of

⁷⁷Although Heaton & Lucas have repeatedly questioned the economic reality of the model.

her utility function. This result has already been hinted at in section 13.1.

Weil (1992) has attempted to resolve the equity premium puzzle by combining a proper risk averse investor with a low probability, highly negatively skewed exogenous consumption risk (“unemployment”) which is independent of stock market risk. His paper includes an example using constant relative risk aversion (which ensures proper risk aversion) in an idealised economy. Predicted asset returns in this case are reasonably close to their observed historical values. This chapter has two main aims. First, a detailed literature review of local proper risk aversion provides insight into the relationship between its various forms and a new type of proper risk aversion emerges naturally. Then a series of counterexamples shows that Weil’s economy is largely driven by assumptions stronger than he explicitly recognises with derivatives higher than the fourth playing an important role. Applying the properties of proper risk aversion more restrictively the equity premium puzzle remains unsolved.

The chapter proceeds as follows. Section 18 provides an integrated approach to local proper risk aversion. As a consequence of this, a new form of proper risk aversion emerges (labelled Basic Risk Aversion) and the concept of “*caution*” is introduced. Section 19 describes Weil’s example, economic interpretation is placed on the fifth derivative of the utility function and the problem being addressed is discussed more fully. Section 20 provides utility functions that are proper risk averse and in many ways similar to power utility and yet estimate very different equity premia for Weil’s example than

power utility. It is inferred that the first four derivatives of the utility function do not fully capture the empirical effects of proper risk aversion. Section 21 concludes.

18 Proper risk aversion

This section provides an integrated approach to local proper risk aversion — a condition under which exogenous income gambles increase the aversion of investors to independent marketable gambles. This section is based in style on that given by Gollier and Pratt (1996) and the associated working paper (Gollier and Pratt (1993)). Nevertheless, the discussion here does develop the existing reviews of this area in a number of ways. By concentrating on local proper risk aversion (proper risk aversion in the presence of small background gambles) it is possible to create an integrated theory that ties together the different strands of proper risk aversion that have been developed by different authors at different times. Also, by concentrating on local proper risk aversion, the intuition is clearer than in the more complicated case of global proper risk aversion (proper risk aversion in the presence of any size background gamble). As a consequence of this development a new type of proper risk aversion — labelled Basic Risk Aversion here — naturally emerges. It is shown that the joint conditions of Basic Risk Aversion and positive third derivative of the utility function (denoted by the term Core Basic Risk Aversion) is a more general class of proper risk aversion than any considered so far. Despite this the necessary and sufficient conditions for

local Core Basic Risk Aversion are the same as the necessary and sufficient conditions for local Standard Risk Aversion. For this reason the properties of decreasing absolute risk aversion and decreasing absolute prudence guarantee proper risk aversion under a broader class of (local) endowment gambles than has previously been recognised by the literature.

18.1 On the utility function

18.1.1 Risk Aversion, Prudence and Temperance

Consider a five times differentiable utility of consumption function $u(\cdot)$ that is monotonic increasing and risk averse across its domain. The following definitions will be useful in the analysis to follow. Consider the utility function at some non-stochastic point of consumption c within the domain of the function. Use the notation $u^{(n)}(c)$ to denote the n^{th} derivative of u at c (so, for example, $u^{(1)}(c) = u'(c)$). Define:

$$\begin{aligned} \mathcal{A} &:= -u^{(2)}(c)/u^{(1)}(c) && \text{Absolute risk aversion} \\ \mathcal{P} &:= -u^{(3)}(c)/u^{(2)}(c) && \text{Absolute prudence} \\ \mathcal{T} &:= -u^{(4)}(c)/u^{(3)}(c) && \text{Temperance} \\ \mathcal{C} &:= -u^{(5)}(c)/u^{(4)}(c) && \text{Caution} \end{aligned}$$

$\mathcal{A}, \mathcal{P}, \mathcal{T}$ are, of course, functions of $u(\cdot)$ and c , but the notation takes this relation as implicit for simplicity. Here absolute risk aversion is attributable to Pratt (1964) and Arrow (1970), prudence to Kimball (1990) and temperance to Kimball (1992) and Eeckhoudt, Gollier and Schlesinger (1993). “Caution” is introduced in this chapter for the first time. These properties can be easily interpreted. Consider $-u^{(1)}(\cdot), u^{(2)}(\cdot), -u^{(3)}(\cdot)$ as “pseudo-utility functions” in their own right. The assumptions of monotonicity and

risk aversion of $u(\cdot)$ does not ensure that $-u^{(1)}(\cdot)$ is risk averse (although it must be increasing) nor that $u^{(2)}(\cdot), -u^{(3)}$ are either increasing or risk averse. Despite this, the measures of risk aversion of $-u^{(1)}, u^{(2)}, -u^{(3)}$ (denoted by $\mathcal{A}_{(-u^{(1)})}, \mathcal{A}_{(u^{(2)})}, \mathcal{A}_{(-u^{(3)})}$ respectively) are defined if we assume that the utility function is five times differentiable and that none of the first four derivatives are equal to zero at c ⁷⁸. By definition $\mathcal{P} = \mathcal{A}_{(-u^{(1)})}$, $\mathcal{T} = \mathcal{A}_{(u^{(2)})}$ and $\mathcal{C} = \mathcal{A}_{(-u^{(3)})}$. Therefore \mathcal{P}, \mathcal{T} and \mathcal{C} can be interpreted as the risk aversion of pseudo-utility functions $-u^{(1)}, u^{(2)}$ and $-u^{(3)}$ respectively⁷⁹. It is easily seen (as noted by Kimball (1990) amongst others) that:

$$\frac{\mathcal{A}'}{\mathcal{A}} = \frac{u^{(3)}}{u^{(2)}} - \frac{u^{(2)}}{u^{(1)}} = -\mathcal{P} + \mathcal{A} = \mathcal{A} - \mathcal{A}_{(-u^{(1)})}$$

Given $\mathcal{A} > 0$ by assumption, the following statements are equivalent: (i) $-u^{(1)}$ is more risk averse than u (ii) $\mathcal{P} > \mathcal{A}$ (iii) the utility function exhibits decreasing absolute risk aversion (DARA) $\mathcal{A}' < 0$. This argument can be taken up one derivative to give:

$$\frac{\mathcal{P}'}{\mathcal{P}} = \frac{u^{(4)}}{u^{(3)}} - \frac{u^{(3)}}{u^{(2)}} = -\mathcal{T} + \mathcal{P} = \mathcal{A}_{(-u^{(1)})} - \mathcal{A}_{(u^{(2)})}$$

The assumptions given for u do not ensure that \mathcal{P} is positive so this time the equivalent statements are: (i) $u^{(2)}$ is more risk averse than $-u^{(1)}$

⁷⁸We will return to occurrences where the higher order derivatives are equal to zero at c as special cases below.

⁷⁹It should be observed that this interpretation of \mathcal{P}, \mathcal{T} is not the only one that provides economic insight. In particular recall that the observation that $\mathcal{A} > 0$ is equivalent to observing that an investor is averse to a mean preserving increase in risk of a marketable gamble. Eeckhoudt, Gollier and Schneider (1995) note that $\mathcal{P}, \mathcal{T} > 0$ can also be readily interpreted in terms of how an investor's aversion to a marketable gamble is influenced by the probability distribution of that gamble.

(ii) $\mathcal{T} > \mathcal{P}$ (iii) \mathcal{P}' and \mathcal{P} are of opposite sign. In particular, for a positive prudence utility function the function must also exhibit decreasing absolute prudence (DAP) $\mathcal{P}' < 0$. Just repeating the argument up one more derivative gives:

$$\frac{\mathcal{T}'}{\mathcal{T}} = \frac{u^{(5)}}{u^{(4)}} - \frac{u^{(4)}}{u^{(3)}} = -\mathcal{C} + \mathcal{P} = \mathcal{A}_{(u^{(2)})} - \mathcal{A}_{(-u^{(3)})}$$

As the sign of \mathcal{T} is again undetermined by the assumptions on u so the equivalent statements are: (i) $-u^{(3)}$ is more risk averse than $u^{(2)}$ (ii) $\mathcal{C} > \mathcal{T}$ (iii) \mathcal{T}' and \mathcal{T} are of opposite sign.

While considering $u, -u^{(1)}, u^{(2)}, -u^{(3)}$ as utility function, it will be helpful to consider the Markowitz risk premia, π, η, ψ, ϕ respectively, for these utility functions for a small gamble \tilde{y} :

$$\begin{aligned} E[u(c + \tilde{y})] &= u(c - \pi) \\ E[-u^{(1)}(c + \tilde{y})] &= -u^{(1)}(c - \eta) \\ E[u^{(2)}(c + \tilde{y})] &= u^{(2)}(c - \psi) \\ E[-u^{(3)}(c + \tilde{y})] &= -u^{(3)}(c - \phi) \end{aligned} \tag{42}$$

If \tilde{y} is sufficiently small so that approximations can be taken in the manner of Pratt (1964):

$$\begin{aligned} \pi &= -E[\tilde{y}] + 0.5\sigma_{\tilde{y}}^2 \mathcal{A} \\ \eta &= -E[\tilde{y}] + 0.5\sigma_{\tilde{y}}^2 \mathcal{P} \\ \psi &= -E[\tilde{y}] + 0.5\sigma_{\tilde{y}}^2 \mathcal{T} \\ \phi &= -E[\tilde{y}] + 0.5\sigma_{\tilde{y}}^2 \mathcal{C} \end{aligned}$$

It is now possible to find relationships between π, η, ψ and ϕ :

$$\begin{aligned} \eta &= \pi + 0.5\sigma_{\tilde{y}}^2(\mathcal{P} - \mathcal{A}) \\ \psi &= \eta + 0.5\sigma_{\tilde{y}}^2(\mathcal{T} - \mathcal{P}) \\ \phi &= \psi + 0.5\sigma_{\tilde{y}}^2(\mathcal{C} - \mathcal{T}) \end{aligned} \tag{43}$$

This confirms the equivalence statements derived above that $-u^{(1)}$ is more risk averse than u if and only if prudence is greater than risk aversion, $u^{(2)}$ is more risk averse than $-u^{(1)}$ if and only if temperance is greater than prudence and $-u^{(3)}$ is more risk averse than $u^{(2)}$ if and only if caution is greater than temperance.

18.1.2 Risk aversion with stochastic consumption

In the previous subsection the Pratt-Arrow measure of absolute risk aversion of utility function u and pseudo-utility functions $-u^{(1)}, u^{(2)}, -u^{(3)}$ was considered at a point of certain consumption c . Now suppose that consumption is stochastic and denoted by $\tilde{c} = c + \tilde{y} + \tilde{e}$. Here c is “certain” income, \tilde{y} is exogenous income risk (not necessarily with zero mean) and \tilde{e} “marketable” risk (that is, comes from trading in a gamble) with zero expectation. To find the absolute risk aversion in the presence of these two sources of risk, a “derived utility function” is used. The derived utility function at c in the presence of background risk \tilde{y} is defined by $E_{\tilde{y}}[u(c + \tilde{y})]$. See Kihlstrom, Romer and Williams (1981), Ross (1981) and Nachman (1982) for a detailed discussion of how the Pratt-Arrow properties of risk aversion transfer to the derived utility function. If \tilde{y} and \tilde{e} are independent so that $E_{\tilde{y}, \tilde{e}} = E_{\tilde{y}} E_{\tilde{e}}$ then the risk premium ϖ for \tilde{e} is defined by:

$$E_{\tilde{y}, \tilde{e}}[u(c + \tilde{y} + \tilde{e})] = E_{\tilde{y}}[u(c + \tilde{y} - \varpi)]$$

Taking Taylor’s series expansions in the usual way of Pratt (1964) gives:

$$\varpi \approx \frac{1}{2}\sigma_{\tilde{e}}^2 \frac{-E_{\tilde{y}}u^{(2)}(c + \tilde{y})}{E_{\tilde{y}}u^{(1)}(c + \tilde{y})} =: \frac{1}{2}\sigma_{\tilde{e}}^2 \mathcal{A}_{\tilde{y}}$$

where $\mathcal{A}_{\tilde{y}}$ denotes absolute risk aversion in the presence of exogenous risk \tilde{y} — that is, the risk aversion of the derived utility function. Notice that $-E[u^{(2)}]/E[u^{(1)}] \neq -E[u^{(2)}/u^{(1)}]$. It is the former and not the latter that gives risk aversion in the presence of background uncertainty. In this section the focus is on small income gambles. In this case it is possible to use the Markowitz risk premia of equation 42 to derive an equation to define the difference between the risk aversion of an investor with stochastic income $c + \tilde{y}$ and one with certain income c :

$$\begin{aligned} \Delta\mathcal{A} := \mathcal{A}_{\tilde{y}} - \mathcal{A} &= -\frac{u^{(2)}(c - \psi)}{u^{(1)}(c - \eta)} + \frac{u^{(2)}(c)}{u^{(1)}(c)} \\ &\approx \psi \frac{u^{(3)}}{u^{(1)}} - \eta \frac{(u^{(2)})^2}{(u^{(1)})^2} \\ &= \psi\mathcal{P}\mathcal{A} - \eta\mathcal{A}^2 \end{aligned}$$

where the second line comes from approximating $u'(c - \eta) \approx u'(c)$ in the denominator. So, by applying equation 43 to the previous equation, for \tilde{y} to cause $\Delta\mathcal{A} > 0$ (under the assumption that $\mathcal{A} > 0$):

$$\pi(\mathcal{A} - \mathcal{P}) < 0.5\sigma_{\tilde{y}}^2[(\mathcal{T} - \mathcal{P})\mathcal{P} + (\mathcal{A} - \mathcal{P})^2] \quad (44)$$

$$\equiv \eta(\mathcal{A} - \mathcal{P}) < 0.5\sigma_{\tilde{y}}^2(\mathcal{T} - \mathcal{P})\mathcal{P} \quad (45)$$

$$\equiv \psi(\mathcal{A} - \mathcal{P}) < 0.5\sigma_{\tilde{y}}^2(\mathcal{T} - \mathcal{P})\mathcal{A} \quad (46)$$

$$\equiv \phi[\mathcal{A} - \mathcal{P}] < 0.5\sigma_{\tilde{y}}^2[(\mathcal{T} - \mathcal{P})\mathcal{A} + (\mathcal{C} - \mathcal{T})(\mathcal{A} - \mathcal{P})] \quad (47)$$

This will provide the basis for much of the discussion that follows.

18.2 Standard and Proper Risk Aversions

This subsection on proper risk aversion is based on that given by Gollier and Pratt (1996) but here includes Very Weak Risk Aversion (defined below) and the concept of Basic Risk Aversion is introduced for the first time. Consider a set of exogenous background gambles $\Sigma_i(c, u)$ for $i = 0, 1, 2, 3, 4a, 4b$. Elements of $\Sigma_i(c, u)$ will be denoted by \tilde{y} . These will be defined as follows:

$$\begin{aligned}\Sigma_0(c, u) &\equiv \{\tilde{y} \mid Eu^{(2)}(c + \tilde{y}) \leq u^{(2)}(c)\}; \\ \Sigma_1(c, u) &\equiv \{\tilde{y} \mid Eu^{(1)}(c + \tilde{y}) \geq u^{(1)}(c)\}; \\ \Sigma_2(c, u) &\equiv \{\tilde{y} \mid Eu(c + \tilde{y}) \leq u(c)\}; \\ \Sigma_3(c, u) &\equiv \{\tilde{y} \mid E\tilde{y} \leq 0\}; \\ \Sigma_{4a}(c, u) &\equiv \{\tilde{y} \mid E\tilde{y} = 0\}; \\ \Sigma_{4b}(c, u) &\equiv \{\tilde{y} \mid \exists y_0 \leq 0 : \tilde{y} = y_0 \text{ with probability } 1\}.\end{aligned}$$

We now define u to be “*proper at c*” with respect to $\Sigma_i(c, u)$ (and the fact that it is proper under Σ_i at c will be denoted by $P_i(c)$) if:

$$Eu(c + \tilde{x} + \tilde{y}) \leq Eu(c + \tilde{y}) \text{ whenever } Eu(c + \tilde{x}) \leq u(c) \text{ and } \tilde{y} \in \Sigma_i(w, u)$$

In words, this can be understood as “Any unattractive gamble cannot be made attractive by the introduction of an independent background gamble in Σ_i if the utility function at c is proper with respect to Σ_i ”. Alternatively it can be said that an investor is at least as risk averse in the presence of $\tilde{y} \in \Sigma_i$ as in its absence if the investor is proper risk averse with respect to Σ_i . That is, $\Delta\mathcal{A} \geq 0$ for $\tilde{y} \in \Sigma_i$ ⁸⁰. We restrict the discussion here to strict

⁸⁰I am grateful to Christian Gollier for the observation that, strictly speaking, this chapter is dealing with “local local proper risk aversion” as “ $\Delta\mathcal{A}(c) \geq 0$ does not imply that $\Delta\mathcal{A}(c') \geq 0$ for $c' \neq c$ ” (private correspondence with author). Therefore this definition of proper risk aversion will only hold if both the marketable and nonmarketable gamble is

| Property | Risk Aversion | Paper |
|-------------|--------------------|-----------------------------|
| $P_0(c)$ | Basic | This chapter |
| $P_1(c)$ | Standard | Kimball (1993) |
| $P_2(c)$ | Proper | Pratt & Zeckhauser (1987) |
| $P_3(c)$ | Risk Vulnerability | Gollier & Pratt (1996) |
| $P_{4a}(c)$ | “Very Weak Proper” | Franke <i>et al.</i> (1995) |
| $P_{4b}(c)$ | DARA | |

Table 3: Definitions of the various forms of Proper Risk Aversion with the initial paper that developed the concept. The term “Very Weak Proper” is used here but Franke *et al.* do not give this property a name. Risk Vulnerability was called Weak Proper Risk Aversion in Gollier and Pratt’s 1993 working paper.

inequality⁸¹.

A utility function is said to be “*proper*” if it is proper at all c within the domain of the utility function. The various concepts of properness that relate to the different forms of $\Sigma_i(c, u)$ are defined in table 3⁸². Property $P_0(c)$ — that of Basic Risk Aversion — is introduced for the first time in this chapter.

This chapter now proceeds by examining necessary and sufficient conditions for these types of proper risk aversion for small exogenous income gambles \tilde{y} . The utility function will then be said to be “*locally proper*” (at c) as opposed to “*globally proper*”, which refers to general endowment gambles in $\Sigma_i(c, u)$. In general, it is not possible to derive simple necessary and

small. This chapter stays with the terminology “local proper risk aversion” for simplicity but it should be borne in mind that it is “local local proper risk aversion” that is being referred to.

⁸¹Except for the alternating sign of derivative condition of Pratt and Zeckhauser (1987) which is a weak inequality condition. This can be seen by looking at exponential utility, which does have alternating sign derivatives but which does not satisfy the strict inequality conditions for proper risk aversion.

⁸²Somewhat ambiguously, “proper risk aversion” and “Proper Risk Aversion” are taken to have different meanings. A utility function is said to be proper risk averse with respect to Σ_i for unspecified i . If $i = 2$ then the investor is Proper Risk Averse.

sufficient conditions for general background gambles. This more limited approach increases the economic intuition and also enable us to consider the inter-relations between the different forms of proper risk aversion.

18.2.1 Local proper risk aversion

The six types of $\Sigma_i(c, u)$ can be divided into two categories. For $i = 3, 4a, 4b$, Σ_i is defined depending on the properties of \tilde{y} alone. For $i = 0, 1, 2$ Σ_i is defined by the relationship between \tilde{y} and u . For the purposes of this subsection, it is useful to consider these two categories of Σ_i separately

$\Sigma_i(c, u)$ for $i = 3, 4a, 4b$

Decompose \tilde{y} into a zero expectation gamble $\tilde{\epsilon}$ and a certain negative shift in income δ . How does this change the absolute risk aversion of the investor with certain background consumption c ? Assume that δ is sufficiently small to ignore terms in δ^2 and above:

$$\begin{aligned} \frac{\Delta \mathcal{A}}{\mathcal{A}} &= [\mathcal{A}]^{-1} \left[\frac{-E[u^{(2)}(c + \tilde{\epsilon} - \delta)]}{E[u^{(1)}(c + \tilde{\epsilon} - \delta)]} \right] - 1 \\ &= \frac{\frac{E[\tilde{\epsilon}^2]}{2} \left(\frac{u^{(4)}}{u^{(2)}} - \frac{u^{(3)}}{u^{(1)}} - \delta \left[\frac{u^{(5)}}{u^{(2)}} - \frac{u^{(4)}}{u^{(1)}} \right] \right) + \delta \left(\frac{u^{(2)}}{u^{(1)}} - \frac{u^{(3)}}{u^{(2)}} \right) + \Omega(\tilde{\epsilon}^3)}{1 - \delta \frac{u^{(2)}}{u^{(1)}} + \Omega(\tilde{\epsilon}^2)} \end{aligned} \quad (48)$$

where the utility function is evaluated at c and where:

$$\begin{aligned}\Omega(\tilde{\epsilon}^3) &= \sum_{n=3}^{\infty} \frac{E[\tilde{\epsilon}^n]}{n!} \left[\left(\frac{u^{(n+2)}}{u^{(2)}} - \frac{u^{(n+1)}}{u^{(1)}} \right) - \delta \left(\frac{u^{(n+3)}}{u^{(2)}} - \frac{u^{(n+2)}}{u^{(1)}} \right) \right] \\ \Omega(\tilde{\epsilon}^2) &= \sum_{n=2}^{\infty} \frac{E[\tilde{\epsilon}^n]}{n!} \left[\frac{u^{(n+1)}}{u^{(1)}} - \delta \frac{u^{(n+2)}}{u^{(1)}} \right]\end{aligned}$$

For small gambles the omega terms, $E[\tilde{\epsilon}^2]\delta$ term and the $\delta u^{(2)}/u^{(1)}$ term in the denominator can be considered sufficiently small to ignore. Therefore, a zero expectation consumption gamble is going to increase the absolute risk aversion of an investor if and only if $\mathcal{P}(\mathcal{T} - \mathcal{A}) = \mathcal{P}(\mathcal{T} - \mathcal{P}) + \mathcal{P}(\mathcal{P} - \mathcal{A}) = -\mathcal{P}' + \mathcal{P}(\mathcal{P} - \mathcal{A}) > 0$ ⁸³. Notice that this does not restrict $\mathcal{P} > 0$ nor $\mathcal{T} > \mathcal{A}$. This condition can be restated to say that $d[u^{(3)}/u^{(1)}]/dc < 0$. This is the necessary and sufficient condition given in Franke, Stapleton and Subrahmanyam (1994) for local Very Weak Proper Risk Aversion⁸⁴. A certain decrease in wealth will increase risk aversion if and only if $\mathcal{P} > \mathcal{A}$ — a condition equivalent to DARA. For local Risk Vulnerability, it is necessary for the utility function to be locally Very Weak Proper and DARA. That is $\mathcal{P} > \mathcal{A}$, $\mathcal{T} > \mathcal{A}$ are necessary conditions for local Risk Vulnerability. These conditions are also sufficient for negative expectation stochastic background risk to increase absolute risk aversion; a condition not covered by either DARA or Very Weak Proper Risk Aversion and yet covered by Risk Vulnerability. Therefore $\mathcal{P} > \mathcal{A}$, $\mathcal{T} > \mathcal{A}$ are necessary and sufficient conditions for local Risk Vulnerability — a result given by Gollier and Pratt (1996).

⁸³If $u^{(3)} = 0$ then \mathcal{T} is not well defined. In this case equation 48 shows that $u^{(4)} < 0$ is necessary and sufficient for the function to be locally Very Weak Proper Risk Averse.

⁸⁴Strictly speaking, Franke *et al.* show that this condition is necessary for an increase in an existing exogenous risk to increase risk aversion. This is slightly different from the comparison of risk aversion in the presence and absence of a background risk. The strict result of Franke *et al.* can also be derived without difficulty from this analysis.

$\Sigma_i(c, u)$ for $i = 0, 1, 2$

The proofs in this subsection are somewhat similar to the ones given in Eeckhoudt and Kimball (1992) and Weil (1992a) where certainty equivalents are used to prove that DAP and DARA are sufficient for global Proper Risk Aversion. There are important differences between their proofs and the one given here, however. They derive a sufficient condition for a global property of one form of proper risk aversion. The results here are, though, both necessary and sufficient and apply to more general classes of proper risk aversion. These conditions do, though, only apply locally. The proofs rely on three straightforward observations: (i) if the first three derivatives are non-zero at c then, for any $i \in \{0, 1, 2\}$, there exists some real number \bar{y} such that, if $\tilde{y} = \bar{y}$ with certainty, then $\tilde{y} \in \Sigma_i$; (ii) For any $\tilde{y} \in \Sigma_i$ the Markowitz risk premium of this gamble on the pseudo-utility function associated with Σ_i will have the opposite sign to the sign of \bar{y} , (iii) By adjusting the expected value of \tilde{y} , it will be possible to create a gamble in Σ_i with Markowitz risk premium on the pseudo-utility function associated with Σ_i arbitrarily close to zero. The proofs now follow:

Return to the Markowitz risk premia given in equation 42. First, suppose that $\tilde{y} \in \Sigma_2(c, u)$. Then, there exists some positive number y_+ such that, if $\tilde{y} = -\bar{y}_+$ with certainty then $\tilde{y} \in \Sigma_2(c, u)$ as $u'(1) > 0$. For this gamble, the right hand side of equation 44 is zero as \tilde{y} has zero variance. This restricts

$\mathcal{A} < \mathcal{P}$ as $\pi > 0$ (π has the opposite sign to $-\bar{y}_+$). Consider the set of gambles in $\Sigma_2(c, u)$ with positive variance, which again is certainly a non-empty subset. The left hand side of equation 44 will always be negative as π must be positive and $\mathcal{A} < \mathcal{P}$ from above. So provided that $(\mathcal{T} - \mathcal{P})\mathcal{P} > -(\mathcal{A} - \mathcal{P})^2$ then the inequality will always hold true. This condition is therefore sufficient for Proper Risk Aversion in the presence of small consumption gambles. It is also necessary as there will be elements of $\Sigma_2(c, u)$ with non-infinitesimal variance and positive expectation with π arbitrarily close to (but always above) zero. For these cases, the right hand side must be strictly positive and so the condition is necessary as well as sufficient. So, necessary and sufficient conditions for local Proper Risk Aversion are $\mathcal{P} > \mathcal{A}$ and $(\mathcal{T} - \mathcal{P})\mathcal{P} > -(\mathcal{A} - \mathcal{P})^2$. These conditions are algebraically equivalent to the local necessary and sufficient conditions given in Pratt and Zeckhauser (1987) ($\mathcal{A}' < 0$, $\mathcal{A}'' > \mathcal{A}'\mathcal{A}$) and Gollier and Pratt (1996) ($\mathcal{P} > \mathcal{A}$, $\mathcal{T}/\mathcal{A} > 2 - (\mathcal{A}/\mathcal{P})$).

Consider now $\Sigma_1(c, u)$. Because $u^{(2)} < 0$ by assumption, there again exists a positive \bar{y}_+ such that $\tilde{y} = -\bar{y}_+ \in \Sigma_1(c, u)$. Given that $\eta > 0$ for this particular risk (opposite sign to $-\bar{y}_+$), applying equation 45 for this \tilde{y} it is clear again that $\mathcal{P} > \mathcal{A}$. Now consider a gamble \tilde{y} with positive variance but mean adjusted so that η is arbitrarily close to zero. Then the right hand side of equation 45 must be positive and this restricts $\mathcal{T} > \mathcal{P}$ (as $\mathcal{P} > \mathcal{A} > 0$). This is also a sufficient condition as the left hand side of equation 45 will always be negative. So $\mathcal{T} > \mathcal{P} > \mathcal{A}$ is necessary and sufficient.

The case of $\Sigma_0(c, u)$ is the most complicated as no assumption is made about the sign of $u^{(3)}$. As $u^{(3)} \neq 0$ by assumption $\tilde{y} = \bar{y} \in \Sigma_0(c, u)$ for positive or negative \bar{y} . First consider the case when $u^{(3)} > 0$. An investor who has positive third derivative of utility and is Basic Risk Averse (at c) will be said to be “*Core Basic Risk Averse*”. In this case $\tilde{y} = -\bar{y}_+ \in \Sigma_0$ so $\psi > 0$ for all gambles in Σ_0 . Therefore, as the left hand side of equation 46 must be negative for this \tilde{y} , this restricts $\mathcal{P} > \mathcal{A}$. Applying the usual arguments now of examining a stochastic gamble \tilde{y} with mean adjusted such that ψ is arbitrarily close to zero, this implies that $\mathcal{T} > \mathcal{P}$ is a necessary condition for local Core Basic Risk Aversion. It is also sufficient through the usual argument. If $u^{(3)} < 0$ the non stochastic gamble in $\Sigma_0(c, u)$ is $\tilde{y} = \bar{y}_+ \in \Sigma_0$, and $\psi < 0$ for all gambles in Σ_0 . As $u^{(3)} < 0$, $\mathcal{P} < 0$ and so the left hand side of equation 46 is certainly negative. Again, by introducing a stochastic gamble with mean adjusted to make ψ arbitrarily close to zero, the usual argument shows that $\mathcal{T} > \mathcal{P}$ is necessary and sufficient. Notice that, in this case, \mathcal{T} is not restricted to be positive. So for Basic Risk Aversion one of the following two sets of conditions are necessary and sufficient: (i) $\mathcal{T} > \mathcal{P} > \mathcal{A}$ (ii) $\mathcal{P} < 0, \mathcal{T} > \mathcal{P}$. The former (Core Basic Risk Aversion) is of more interest here. Notice that $u^{(3)}$ is bounded away from zero from above in this case. So if a utility function is Basic Risk Averse over its domain and Core Basic Risk Averse at some point then it is Core Basic Risk Averse over its domain⁸⁵. As the focus here is on Core Basic Risk Aversion the restriction that $u^{(3)} \neq 0$ is

⁸⁵Provided that the third derivative is continuous, which is assured as the utility function is five times differentiable.

not constraining.

Result 1 *A necessary and sufficient condition for Standard and Core Basic Risk Aversion in the presence of small consumption gambles is:*

$$\mathcal{T} > \mathcal{P} > \mathcal{A}$$

While Kimball (1993) has shown that DAP and DARA are necessary and sufficient for global Standard Risk Aversion the author believes that this is the first time that it has been shown that these conditions are also necessary for local Standard Risk Aversion. These conditions are also necessary and sufficient conditions for local Core Basic Risk Aversion.

Examine the inter-relations between the various forms of proper risk aversion. If an investor is Core Basic Risk Averse then it is necessary that her temperance is greater than her prudence (this is true for local Core Basic Risk Aversion and so must also be true for global Core Basic Risk Aversion). This implies that $u^{(2)}$ is more risk averse than $-u^{(1)}$ and so $\Sigma_1(c_u) \subset \Sigma_0(c_u)$. So if an investor is Core Basic Risk Averse then she is certainly Standard Risk Averse. If an investor is Standard Risk Averse then her prudence is greater than her risk aversion. This implies that $-u^{(1)}$ is more risk averse than u and so $\Sigma_2(c_u) \subset \Sigma_1(c_u)$. So if an investor is Standard Risk Averse then she is certainly Proper Risk Averse. That $\Sigma_3 \subset \Sigma_2$ follows from the risk aversion of u . So Proper Risk Aversion is sufficient for Risk Vulnerability. Finally,

by definition and the assumption that $u^{(1)} > 0$, $\Sigma_{4a} \subset \Sigma_3$ and $\Sigma_{4b} \subset \Sigma_3$. So Risk Vulnerability is sufficient for Very Weak Proper Risk Aversion and decreasing absolute risk aversion. Notice that local Core Basic Risk Aversion is a strictly stronger property than local Standard Risk Aversion — that is, $\Sigma_1(c, u)$ is a strict subset of $\Sigma_0(c, u)$. Despite this, the necessary and sufficient conditions for local Core Basic and Standard Risk Aversions are identical. Hence the conditions given by Kimball (1993) for Standard Risk Aversion apply locally to a greater class of consumption gambles than he recognises.

Is it possible to broaden the class of local consumption gambles \tilde{y} further and still have DAP and DARA as necessary and sufficient conditions to have the utility function proper risk averse with respect to this set of gambles? Define $\Sigma_{-1}(c, u) = \{\tilde{y} | E[u^{(3)}(c + \tilde{y})] \geq u^{(3)}(c)\}$. Assume that the utility function is DAP and DARA and five times differentiable. Do these conditions ensure that the utility function is proper with respect to Σ_{-1} ? Because DAP ensures that $u^{(4)} < 0$ this means that $\tilde{y} = -\bar{y}_+ \in \Sigma_{-1}$. So, $\phi > 0$ for all gambles in Σ_{-1} . As $\mathcal{P} > \mathcal{A}$ this implies that the left hand side of equation 47 is certainly negative but can be made arbitrarily close to zero as usual by adjusting the mean of \tilde{y} . So, the right hand side must be positive. The assumption of DAP and DARA says nothing about the relationship between \mathcal{C} and \mathcal{T} . Notice, though, that this form of proper risk aversion will only be of interest if $\Sigma_0 \subset \Sigma_{-1}$ or $-u^{(3)}$ is more risk averse than $u^{(2)}$. This only holds if $\mathcal{C} > \mathcal{T}$, which is the case now considered. Considering the right hand side of equation 47 with $\mathcal{C} > \mathcal{T}$ and $\mathcal{A} < \mathcal{P}$ the condition that $\mathcal{T} > \mathcal{P}$ (DAP) is

no longer sufficient to ensure that the right hand side is positive. So DAP and DARA do not imply local properness with respect to Σ_{-1} which is the next obvious progression from Core Basic Risk Aversion.

18.2.2 Global proper risk aversion

The chapter so far has focused on local proper risk aversion; that is, proper risk aversion in the presence of small background gambles that are independent of marketable gambles. However, in general, even if a utility function is locally proper with respect to $\Sigma_i(c, u)$ for some i across its domain this is no guarantee that the utility is globally proper with respect to i . Authors have, in general, had great difficulty in establishing tractable conditions that are necessary and sufficient for global proper risk aversion. The only simple necessary and sufficient condition that has been established for global proper risk aversion is attributable to Kimball (1993). He showed that, if a utility function exhibited DAP and DARA across its domain, then this is necessary and sufficient for global Standard Risk Aversion. Therefore, in this case, if the local conditions hold across the domain, then this is necessary and sufficient for the global condition to hold.

For Proper Risk Aversion, the analogous result does not hold. Gollier and Pratt (1996) show that if a utility function is locally Proper Risk Averse across its domain then this is sufficient for the function to be globally Risk Vulnerable. The sufficiency condition for global Proper Risk Aversion originally developed by Pratt and Zeckhauser (1987) of alternating sign deriva-

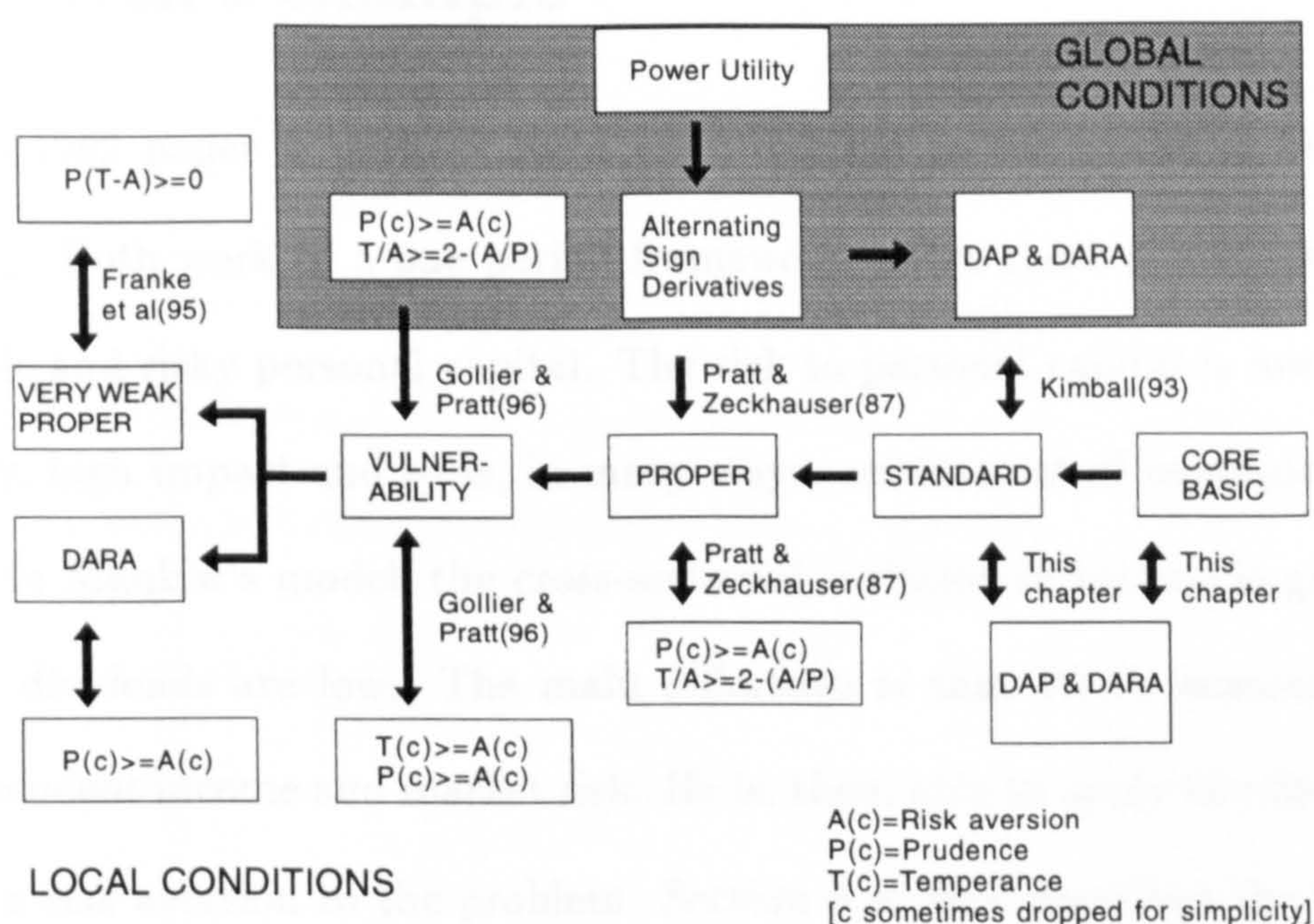


Figure 7: Sufficient and necessary & sufficient conditions for the various forms of local and global proper risk aversion.

tives of the utility function is, in fact, more restrictive than the condition of DAP and DARA that ensure the stronger property of global Standard Risk Aversion. That is, there are no known easily tractable restrictions on the utility function that are sufficient for global Proper Risk Aversion and yet are weaker than necessary for global Standard Risk Aversion.

Despite having introduced the notion, the author makes no attempt here to develop tractable sufficient conditions for global Basic Risk Aversion. This is seen as an area for possible future development. Figure 7 represents the interrelationships between the various forms of local and global proper risk aversion.

19 Weil's example

Weil's 1992 paper is perhaps most easily related to the model of Mankiw (1986). Both work in a one period framework where there is a marketable gamble and risky personal capital. The risk to personal capital is low probability, high impact and so is, in many ways, reminiscent of unemployment risk. In Mankiw's model, the cross-sectional variation in income is greatest when dividends are low. The main difference is that Weil's economy has independent income and market risk. He is, then, able to apply the theory of proper risk aversion to the problem. Section 5 of Weil describes the following economy. At time 0 there are two assets that can be traded, a riskfree bond and a share. At $t = 1$, each investor stands a probability $(0.01, 0.99)$ of receiving income $y = (1, 3)$. The bond will pay 1 with certainty and the share will pay a dividend $d = (0.7, 1)$ with probability $(0.1, 0.9)$. These two risks are independent. It is assumed that utility is $U(c) = c^{1-\gamma}/(1-\gamma)$. Consumption c at time 1 will be $y + d$. The predicted equity premium and riskfree rate in this economy is compared with an identical economy where the probability of unemployment ($y = 1$) is zero. It is shown that introducing the risk of unemployment reduces the required value of γ to make the equity premium 6% from c.17 to c.8. For values of γ around 15, it is shown that the predicted equity premium is increased around threefold by the possibility of unemployment. The impact when unemployed income is zero is even more pronounced. The example also shows a dramatic impact on the predicted riskfree rate through the precautionary savings motive.

While, in Weil's example, the exogenous background risk has small negative expectation (as the baseline case has $y = 3$ with certainty), the change in expectation of consumption caused by this risk is a second order effect. For the remainder of this section it is convenient to assume that Weil's example has zero expectation and apply the theory of Risk Vulnerability. The following result will also prove useful:

Result 2 *The relative change in absolute risk aversion caused to a Risk Vulnerable investor by a small zero expectation uninsurable consumption background risk is decreasing in expected consumption if and only if:*

$$\mathcal{C} > \mathcal{P} + \mathcal{A} \left[1 - \frac{\mathcal{A}}{\mathcal{T}} \right]$$

Proof. From equation 48 $d/dc[\Delta\mathcal{A}/\mathcal{A}]|_{\delta=0} = d/dc[(U^{(4)}/U^{(2)}) - (U^{(3)}/U^{(1)})] < 0$. By calculating the differential and remembering that $\mathcal{PT} > 0$, result 2 follows direct.

It is not possible to establish whether result 2 or decreasing temperance (which implies and is implied by $\mathcal{C} > \mathcal{T}$ for a positive temperance utility function) imposes a stronger restriction on \mathcal{C} . This is because Risk Vulnerability places no restrictions on the relationship between \mathcal{T} and \mathcal{P} . Even if the requirement for Standard Risk Aversion and local Core Basic Risk Aversion

is imposed ($\mathcal{T} > \mathcal{P}$), the stronger of these conditions can vary. To see this, try $\mathcal{A} = 1, \mathcal{P} = 2$ and then $\mathcal{T} = 2.1$ or 3 . Decreasing temperance requires $\mathcal{C} > \mathcal{T} > \mathcal{P}$. So $\mathcal{C} > 2.1, 3$ respectively. Result 2 requires $\mathcal{C} > 2.52, 2.66$ respectively. Therefore decreasing temperance is neither necessary nor sufficient for a decrease in expected consumption to lead to a greater relative increase in risk aversion from the introduction of the same exogenous consumption risk⁸⁶. Notice also that result 2 need not hold for a Very Weak Proper Risk Averse investor even though the income risk is assumed to have zero expectation. This is because \mathcal{PT} need not be positive in this case.

Power utility is proper risk averse under all the definitions given in this chapter. Further, it has fifth and sixth derivatives that will ensure that $\Delta\mathcal{A}/\mathcal{A}$ is monotonic decreasing and convex in c . So, examples that use power utility may be exploiting proper risk aversion phenomena or may be appealing to properties of high order derivatives (greater than four) of the utility function. Result 2 also places, for the first time, intuition on the role that the fifth derivative of the utility function plays. It informs us of how the change in aversion of a proper risk averse investor to a marketable gamble on the introduction of a small zero expectation background exogenous consumption gamble is influenced by the expected consumption level of the investor.

⁸⁶It should be noted that if $\Delta\mathcal{A}/\mathcal{A}$ is required to be convex in c , as well as monotonic decreasing, then this will place economic interpretation on the sixth derivative of the utility function. The algebraic restrictions on $u^{(6)}(c)$ that will ensure this convexity is too messy to present here but can be easily established by the interested reader.

19.1 The issue to be addressed

If an investor is globally Risk Vulnerable then, by definition, the predicted equity premium in the presence of Weil's unemployment risk will be higher than in its absence. Standard Risk Aversion implies Risk Vulnerability and it is known that, for global Standard Risk Aversion, restrictions must be placed on the first four derivatives of the utility function but place no limitations on higher order derivatives. To quote two leading authors:

To tell into which direction one is misled by the representative agent assumption when there is idiosyncratic labor income risk, it is necessary to know how consumers' attitudes towards dividend risk are affected by the existence of background uninsurable labor income risk — an aspect of behavior determined by the signs and magnitudes of higher order (i.e., larger than 3) derivatives of the utility function.

While it is difficult to get introspective knowledge of those high derivatives, it is fortunate that all commonly used utility functions (except for the watershed case of exponential utility, and for quadratic utility) guarantee — because they exhibit DARA and DAP — that the representative agent model underpredicts, and sometimes by a very large factor, the size of the equity premium.

(Weil (1992) p.788)

... the combination of monotonicity, concavity, decreasing absolute risk aversion and decreasing absolute prudence imposes no requirements on the fifth and higher derivatives of the utility function...

(Kimball (1993) p.599)

The remainder of this chapter emphasises an important issue that arise from these quotations. In the first paragraph, Weil appreciates that it is all higher order derivatives that drive the theoretical equity premium. In the second paragraph it is implied that it is the properties of DAP and DARA that cause the “very large factor” underprediction in the representative agent

model. Similarly Kimball implies that the impact of background risk is well captured by the first four derivatives of the utility function. Here it is argued that, while DAP and DARA cause a *theoretical* difference to the predicted equity premium, the *magnitude* of the predicted equity premium in the example of Weil (1992a) is driven by higher order derivatives of the utility function.

It is known from above that $\mathcal{A}_{\tilde{y}} = -E_{\tilde{y}}[u^{(2)}(c + \tilde{y})]/E_{\tilde{y}}[u^{(1)}(c + \tilde{y})]$. While the consumption gamble in Weil's example has low variance it is highly negatively skewed and spans a wide domain. How does one estimate $\mathcal{A}_{\tilde{y}}$? There are two potential solutions. First, one can explicitly estimate $u^{(2)}, u^{(1)}$ at all points over the consumption domain and the expectations can be calculated directly. This is not the solution that Weil takes. The alternative is to assume that the utility function can be expressed as an equation (or set of equations) over the consumption domain and the expectation can be calculated with respect to this (these) equation(s). This is what Weil does — he assumes that the utility function can be described by power form with a fixed parameter of relative risk aversion over the total consumption domain. The question then arises as to how to estimate the equation(s) that describe the investor's utility. Because the probability of the low consumption state is so small financial economists are likely to estimate the equation describing the utility function close to the point of expected consumption. Over a small local interval there are a number of equations that will appear almost indistinguishable. However over the large consumption domain covered by Weil's

example they might appear very different. So, what degree of accuracy does one have to achieve locally in estimating the equation describing the utility function in order to have robustness in estimating the equity premium in Weil's example? Unlike chapter 3, the assumption that utility is additively time separable is not relaxed, nor is it assumed that different investors have different utility functions. The analysis and quotations presented above might suggest that all utility functions that are locally similar to the first four derivatives and that are both DAP and DARA over the whole domain will predict similar equity premia for Weil's example. This chapter shows that this is not so. So, the focus of this chapter is on the functional form of the utility function and not parameter estimation.

Why should derivatives higher than the fourth matter? From result 2 it is clear that the fifth derivative tells us about how $\Delta\mathcal{A}$ changes as an investor's expected consumption changes. Given that Weil's example is over a large consumption domain this might imply that, for different levels of caution, the impact of this highly negatively skewed gamble might have a different quantitative effect. The sixth derivative tells us about the convexity of $\Delta\mathcal{A}$, which again might be of empirical significance. Therefore it is not clear why we should assume that the first four derivatives capture the important quantitative as well as qualitative effects for this example. Considering this issue more formally, examine the accuracy of the Taylor's series expansion given in equation 48 for an example similar to that of Weil with $\delta = 0$. As with Weil's example it is assumed in this section that consumption can take

one of two levels c_1, c_2 with $c_1 < c$ without loss of generality and where c is the level of expected consumption. It is assumed that c_1 will occur with probability p , $1-p$ for c_2 . Due to the form of equation 48 it is known that the volatility of consumption σ_c^2 influences $\Delta \mathcal{A}$ and so this will be kept constant throughout. Therefore, given that the expectation of $\tilde{\epsilon}$ must be zero, this gives only one degree of freedom in the choice of p, c_1, c_2 . Without loss of generality, it will be assumed that c_1 is allowed to vary. In this case:

$$\begin{aligned} p &= \sigma_c^2 / [(c - c_1)^2 + \sigma_c^2] \\ c_2 &= [c - pc_1] / [1 - p] \end{aligned}$$

Rearranging:

$$E[\tilde{\epsilon}^n] = p \left[(-1)^n + \left(\frac{p}{1-p} \right)^{n-1} \right] (c - c_1)^n$$

Assume that the utility function takes power form with coefficient of relative risk aversion γ . It is clear that the term involving $E[\tilde{\epsilon}^n]$ in $\Omega(\tilde{\epsilon}^3)$ is given by

$$\text{Term in } E[\tilde{\epsilon}^n] = (-1)^n \frac{(\gamma + 1)(\gamma + 2) \dots (\gamma + n - 1) c^{-n}}{(n - 1)!} E[\tilde{\epsilon}^n]$$

Using the simple two level consumption model presented here, this can be rearranged to give:

$$\begin{aligned} \text{Term in } E[\tilde{\epsilon}^n] &= \frac{(\gamma + 1) \dots (\gamma + n - 1)}{(n - 1)!} \left[1 - \frac{c_1}{c} \right]^n p \left[1 + (-1)^n \left(\frac{p}{1-p} \right)^{n-1} \right] \\ &\approx \frac{(\gamma + 1) \dots (\gamma + n - 1)}{(n - 1)!} \left[1 - \frac{c_1}{c} \right]^n p \end{aligned}$$

where the approximation works for highly negatively skewed gambles where p is very small. Taking ratios:

$$\frac{\text{Term in } E[\tilde{\epsilon}^{n+1}]}{\text{Term in } E[\tilde{\epsilon}^n]} \approx \frac{\gamma + n}{n} \left[1 - \frac{c_1}{c} \right]$$

It is now clear that the term in $E[\tilde{\epsilon}^{n+1}]$ is less than the term in $E[\tilde{\epsilon}^n]$ if and only if $n > \gamma[(c/c_1) - 1]$. This puts a lower bound on the possible number of Taylor's series terms that must be used to get a local approximation. Of course, the true number of required terms must be higher than this as this value of n gives the greatest individual term in the Taylor's expansion. So, if $c_1 = 0.4c$ and $\gamma = 10$, an example similar to the situation under which Weil (1992a) can explain the equity premium, the terms in the Taylor's series expansion start diminishing at the *sixteenth* term. It is therefore suggested here that any local approximation of the Taylor's series expansion to provide a reasonable indication of the magnitude of the effect of proper risk aversion caused by a small background consumption gamble is dependent on very high order derivatives of the utility function.

It should, though, also be noted that terms in $E[\tilde{\epsilon}^n]$ in $\Omega(\tilde{\epsilon}^2)$ also decay very slowly and all are positive⁸⁷. Therefore, by terminating the Taylor's series expansion at the fourth derivative creates two offsetting truncation errors. The numerator and denominator of equation 48 are both underestimated and so the relative approximation given by multiplying the variance of

⁸⁷It can easily be verified that, for $\Omega(\tilde{\epsilon}^2)$, $(\text{Term in } E[\tilde{\epsilon}^{n+1}]) / (\text{Term in } E[\tilde{\epsilon}^n]) \approx [(\gamma + n)/(n + 1)](1 - c_1/c)$. For $\gamma = 10$, $c_1 = 0.4c$, the terms start to diminish at around the 13th term.

the consumption gamble by a term involving the first to fourth derivative of the utility function will depend on how these errors offset. The conclusion of this chapter is that if idiosyncratic labour income risk is included when modelling an economy and it is assumed that the utility function can be described over the whole consumption domain by a simple equation with fixed parameters the predicted equilibrium asset prices can vary substantially depending on the influence of very high order derivatives of the equation. Examples in the next section demonstrate this point.

20 The magnitude of the equity premium

20.1 Creating alternate utility functions

To show the importance of the higher order terms, this chapter applies three alternative utility functions to the example of Weil (1992a). Two have “constant DAP”. By this, it is meant that $\mathcal{P}' = -k$ for some positive constant k . These two utilities functions will differ by the level to which k is set. The third has “constant AAP” where AAP stands for Accelerating Absolute Prudence. Here constant AAP means that $\mathcal{P}'' = -l$ for some positive constant l . A fuller description follows:

- “To $U^{(3)}$ ” utility. This utility function will be constant DAP. The value of k will be set very close to zero ($k = 1E^{-5}$). The first three derivatives of the utility function will be set to be equal to those of power utility at the point of expected consumption.

- “To $U^{(4)}$ ” utility. This utility function will be constant DAP. The value of k will be set so that the utility function has the same rate of change of absolute prudence at the point of expected future consumption as power utility. The first three derivatives of the utility function will also be set to be equal to those of power utility at the point of expected consumption.
- “To $U^{(5)}$ ” utility. This utility function will be constant AAP. The value of l will be set so that the utility function has the same acceleration of absolute prudence at the point of expected future consumption as power utility. The first four derivatives of the utility function will also be set to be equal to those of power utility at the point of expected consumption.

It is possible to find analytical form for $U^{(2)}$ for constant DAP and constant AAP utility functions. $\mathcal{P} = -U^{(3)}/U^{(2)}$, so for constant DAP, $U^{(3)}(c) = k(c + b)U^{(2)}(c)$ for some constant b . Similarly for constant AAP $U^{(3)}(c) = (l/6)(3c^2 + 2mc + n)U^{(2)}(c)$ for constants m, n . Solving:

$$\begin{aligned} \text{Constant DAP } U^{(2)}(c) &= -\text{Exp} \left[\frac{k}{2}(c + b)^2 - a \right] \\ \text{Constant AAP } U^{(2)}(c) &= -\text{Exp} \left[\frac{l}{6}(c^3 + mc^2 + nc + q) \right] \end{aligned}$$

where a, q are constants of integration. It is not possible to construct analytical form for $U^{(1)}$. The next issue is to resolve the values of a, b, k, l, m, n, q . The aim is to have the utility functions looking identical to power utility with

coefficient of relative risk aversion γ at some point C (the value of next period's expected consumption) up to the third, fourth or fifth derivative. For constant DAP, it will be demanded that the constant DAP utility function will have the same first (the starting point for the Taylor's series expansion), second (which will determine a) and third (which, as a ratio with the second derivative, will determine b) derivatives as the power utility at $c = C$. For these conditions to be satisfied:

$$\begin{aligned} U^{(1)}(C) &= C^{-\gamma} \\ b &= -(1 + \gamma + C^2 k)/(Ck) \\ a &= k(C + b)^2/2 + (\gamma + 1)\ln(C) - \ln(\gamma) \end{aligned}$$

Next determine k . By definition, $k = 1E^{-5}$ for “to $U^{(3)}$ ” utility. For “to $U^{(4)}$ ” utility we also require the utility function to have same coefficient of \mathcal{P}' at C . In this case $k = (\gamma + 1)/C^2$, $b = -2C$ and $a = (\gamma + 1)/2 + (\gamma + 1)\ln(C) - \ln(\gamma)$. Solving for “to $U^{(5)}$ ” utility, l is determined by the acceleration of absolute prudence, m is solved by the decrease in absolute prudence, n is solved by the level of absolute prudence and q is solved by the second derivative of the utility function. These give:

$$\begin{aligned} U^{(1)}(C) &= C^{-\gamma} \\ l &= -2(\gamma + 1)/C^3 \\ m &= -9C/2 \\ n &= 9C^2 \\ q &= C^3 \left[\frac{-3\ln(\gamma)}{\gamma + 1} + 3\ln(C) - 5.5 \right] \end{aligned}$$

Values for the second, third and four derivatives of the utility function can now be calculated analytically at all points along the domain. Using these three derivatives, the value of marginal utility for all three of these functions

can be determined numerically using a Taylor's series expansion through a Turbo Pascal programme. The step size was $1/10,000$ and a sensitivity analysis was undertaken to ensure that the results were not sensitive to this step size⁸⁸. It was also checked that the utility functions were well behaved across the domain. That is, it was required that the signs of the first four derivatives of each utility function alternate (the first being positive) and that each have DAP and DARA at all points on the domain. All three utility functions were well defined in this sense. Values of the first four derivatives of these utility functions are given at seven consumption points in the range $[1, 4]$ for $\gamma = 3, 15$ in tables 4 and 5 respectively.

⁸⁸Graphs were also produced for $1/500$ step size that the author cannot distinguish from those presented here. In order to check the sensitivity, the marginal utility of power form was estimated using the same Taylor's series technique. For $\gamma = 15$ at $c = 1$, $U^{(1)}$ was estimated at 1.0015 against its true value of 1.

| Consump | U' | U'' | U''' | $U^{(4)}$ | A | A' | P | P' |
|---|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| To $U^{(3)}$ utility (small constant DAP) | | | | | | | | |
| 1 | 0.2454402 | -0.244421 | 0.2475217 | -0.250665 | 0.9958457 | -0.016772 | 1.0126877 | -1E-05 |
| 1.5 | 0.1495376 | -0.147311 | 0.1491796 | -0.151073 | 0.9851117 | -0.027161 | 1.0126827 | -1E-05 |
| 2 | 0.0917375 | -0.088784 | 0.0899097 | -0.09105 | 0.9678063 | -0.043427 | 1.0126777 | -1E-05 |
| 2.5 | 0.0569014 | -0.05351 | 0.0541882 | -0.054875 | 0.9403995 | -0.067966 | 1.0126727 | -1E-05 |
| 3 | 0.0359058 | -0.032251 | 0.0326591 | -0.033073 | 0.8981994 | -0.102815 | 1.0126677 | -1E-05 |
| 3.5 | 0.0232517 | -0.019437 | 0.0196836 | -0.019933 | 0.8359601 | -0.147716 | 1.0126627 | -1E-05 |
| 4 | 0.0156239 | -0.011715 | 0.0118633 | -0.012014 | 0.7498149 | -0.197083 | 1.0126577 | -1E-05 |
| To $U^{(4)}$ utility (locally calculated DAP) | | | | | | | | |
| 1 | 0.4725475 | -0.745737 | 1.31917 | -2.524726 | 1.5781212 | -0.301147 | 1.7689473 | -0.256369 |
| 1.5 | 0.2229542 | -0.317966 | 0.5217075 | -0.937515 | 1.4261516 | -0.306068 | 1.6407627 | -0.256369 |
| 2 | 0.1135434 | -0.144548 | 0.2186399 | -0.367768 | 1.2730627 | -0.304918 | 1.5125781 | -0.256369 |
| 2.5 | 0.0623909 | -0.070061 | 0.0969921 | -0.152237 | 1.1229372 | -0.293599 | 1.3843935 | -0.256369 |
| 3 | 0.0368843 | -0.036206 | 0.045482 | -0.066417 | 0.9816023 | -0.269555 | 1.2562089 | -0.256369 |
| 3.5 | 0.0233193 | -0.019949 | 0.0225025 | -0.030498 | 0.8554546 | -0.233171 | 1.1280244 | -0.256369 |
| 4 | 0.015625 | -0.011719 | 0.0117169 | -0.014719 | 0.750002 | -0.187379 | 0.9998398 | -0.256369 |
| To $U^{(5)}$ utility (locally calculated AAP) | | | | | | | | |
| 1 | 0.6517192 | -1.299563 | 3.0328815 | -7.908861 | 1.9940532 | -0.677415 | 2.3337708 | -0.6393 |
| 1.5 | 0.2607591 | -0.437076 | 0.8874147 | -2.052814 | 1.6761659 | -0.593665 | 2.0303465 | -0.574397 |
| 2 | 0.1209892 | -0.169699 | 0.2985645 | -0.611747 | 1.402599 | -0.500412 | 1.759374 | -0.509493 |
| 2.5 | 0.0636129 | -0.074838 | 0.1138173 | -0.206372 | 1.1764562 | -0.405168 | 1.5208533 | -0.44459 |
| 3 | 0.0370161 | -0.036884 | 0.0484939 | -0.077763 | 0.9964183 | -0.317226 | 1.3147844 | -0.379686 |
| 3.5 | 0.0233241 | -0.019988 | 0.0228096 | -0.032321 | 0.8569669 | -0.24355 | 1.1411673 | -0.314782 |
| 4 | 0.015625 | -0.011719 | 0.0117188 | -0.014647 | 0.75 | -0.187502 | 1.000002 | -0.249879 |
| Power utility | | | | | | | | |
| 1 | 1 | -3 | 12 | -60 | 3 | -3 | 4 | -4 |
| 1.5 | 0.2962963 | -0.592593 | 1.5802469 | -5.26749 | 2 | -1.333333 | 2.6666667 | -1.777778 |
| 2 | 0.125 | -0.1875 | 0.375 | -0.9375 | 1.5 | -0.75 | 2 | -1 |
| 2.5 | 0.064 | -0.0768 | 0.12288 | -0.24576 | 1.2 | -0.48 | 1.6 | -0.64 |
| 3 | 0.037037 | -0.037037 | 0.0493827 | -0.082305 | 1 | -0.333333 | 1.3333333 | -0.444444 |
| 3.5 | 0.0233236 | -0.019992 | 0.0228476 | -0.032639 | 0.8571429 | -0.244898 | 1.1428571 | -0.326531 |
| 4 | 0.015625 | -0.011719 | 0.0117188 | -0.014648 | 0.75 | -0.1875 | 1 | -0.25 |

Table 4: Properties of the first four derivatives of the four utility functions used in this chapter at seven values in the range [1,4] with the coefficient of relative risk aversion equal to 3 at the point of next period's expected consumption in each case.

From looking at these tables, the key insight that lies behind this chapter becomes clear. This table was created with C set at 3.95 (this is average consumption when unemployed income is 1/3 employed income). Take $\gamma = 15$, which is close to the point in Weil's example where the introduction of unemployment risk makes the largest change to the predicted equity premium. At $c = 3$, which is 76% of expected consumption, $\mathcal{A} = 5, 4.955, 4.800, 4.043$ for power utility "to $U^{(5)}$ ", "to $U^{(4)}$ " and "to $U^{(3)}$ " utility respectively. Given the low precision in estimates of the coefficient of risk aversion of individuals (as highlighted in chapter 3), it is highly unlikely that a financial economist could determine which of these functions best reflects investor preferences over the range $[3, 3.95]$. At $c = 1$, $\mathcal{A} = 15, 9.042, 6.924, 4.049$ for the four utility functions. The distinctions between the functions are much clearer at this point. Therefore, if examples are being driven by values of \mathcal{A} at points some way distant from C , and given that it is very difficult to precisely estimate the functional form of investor preferences locally, the quantitative effects of the introduction of unemployment cannot be established with precision.

20.2 Results and discussion

The results are presented in figures 8 and 9. These figures give the ratio of the predicted equity premium or riskfree rate in the presence of unemployment risk to the predicted value in the absence of unemployment risk for various levels of risk aversion. For figures 8, 9 unemployment income is 1 and 0 respectively. The x-axes give the coefficients of relative risk aversion (Gamma) of the power utility that the three alternate utility functions

| Consump | U' | U'' | U''' | U'''' | A | A' | P | P' |
|---|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| To $U^{(3)}$ utility (small constant DAP) | | | | | | | | |
| 1 | 0.0001632 | -0.000661 | 0.0026769 | -0.010843 | 4.0490178 | -0.006659 | 4.0506624 | -1E-05 |
| 1.5 | 2.154E-05 | -8.72E-05 | 0.0003532 | -0.001431 | 4.0490014 | -0.006705 | 4.0506574 | -1E-05 |
| 2 | 2.842E-06 | -1.15E-05 | 4.661E-05 | -0.000189 | 4.0489099 | -0.007055 | 4.0506524 | -1E-05 |
| 2.5 | 3.75E-07 | -1.52E-06 | 6.15E-06 | -2.49E-05 | 4.0482502 | -0.009705 | 4.0506474 | -1E-05 |
| 3 | 4.955E-08 | -2E-07 | 8.115E-07 | -3.29E-06 | 4.0432899 | -0.029728 | 4.0506424 | -1E-05 |
| 3.5 | 6.598E-09 | -2.64E-08 | 1.071E-07 | -4.34E-07 | 4.0061219 | -0.178335 | 4.0506374 | -1E-05 |
| 4 | 9.311E-10 | -3.49E-09 | 1.413E-08 | -5.72E-08 | 3.7463254 | -1.140033 | 4.0506324 | -1E-05 |
| To $U^{(4)}$ utility (locally calculated DAP) | | | | | | | | |
| 1 | 0.0082714 | -0.057273 | 0.4052514 | -2.926208 | 6.9242005 | -1.04963 | 7.0757891 | -1.025477 |
| 1.5 | 0.0002958 | -0.001893 | 0.0124233 | -0.083476 | 6.3982691 | -1.054317 | 6.5630508 | -1.025477 |
| 2 | 1.377E-05 | -8.08E-05 | 0.0004891 | -0.003042 | 5.8696101 | -1.060652 | 6.0503124 | -1.025477 |
| 2.5 | 8.36E-07 | -4.46E-06 | 2.471E-05 | -0.000141 | 5.3372236 | -1.069315 | 5.5375741 | -1.025477 |
| 3 | 6.629E-08 | -3.18E-07 | 1.599E-06 | -8.36E-06 | 4.800075 | -1.078869 | 5.0248358 | -1.025477 |
| 3.5 | 6.882E-09 | -2.93E-08 | 1.323E-07 | -6.27E-07 | 4.2611057 | -1.069502 | 4.5120974 | -1.025477 |
| 4 | 9.313E-10 | -3.49E-09 | 1.397E-08 | -5.94E-08 | 3.7500421 | -0.934949 | 3.9993591 | -1.025477 |
| To $U^{(5)}$ utility (locally calculated AAP) | | | | | | | | |
| 1 | 0.0584176 | -0.528196 | 4.930755 | -47.37971 | 9.0417346 | -2.652379 | 9.335083 | -2.557201 |
| 1.5 | 0.0008643 | -0.006758 | 0.0548861 | -0.461279 | 7.8190873 | -2.3637 | 8.1213859 | -2.297587 |
| 2 | 2.283E-05 | -0.000154 | 0.0010808 | -0.007919 | 6.7284147 | -2.079627 | 7.037496 | -2.037973 |
| 2.5 | 1.006E-06 | -5.81E-06 | 3.534E-05 | -0.000225 | 5.7724881 | -1.794812 | 6.0834133 | -1.778358 |
| 3 | 6.916E-08 | -3.43E-07 | 1.802E-06 | -1E-05 | 4.9554875 | -1.504735 | 5.2591377 | -1.518744 |
| 3.5 | 6.903E-09 | -2.96E-08 | 1.349E-07 | -6.53E-07 | 4.281862 | -1.210942 | 4.5646693 | -1.25913 |
| 4 | 9.313E-10 | -3.49E-09 | 1.397E-08 | -5.94E-08 | 3.7499996 | -0.937532 | 4.0000081 | -0.999515 |
| Power utility | | | | | | | | |
| 1 | 1 | -15 | 240 | -4080 | 15 | -15 | 16 | -16 |
| 1.5 | 0.0022837 | -0.022837 | 0.2435902 | -2.760689 | 10 | -6.666667 | 10.666667 | -7.111111 |
| 2 | 3.052E-05 | -0.000229 | 0.0018311 | -0.015564 | 7.5 | -3.75 | 8 | -4 |
| 2.5 | 1.074E-06 | -6.44E-06 | 4.123E-05 | -0.00028 | 6 | -2.4 | 6.4 | -2.56 |
| 3 | 6.969E-08 | -3.48E-07 | 1.858E-06 | -1.05E-05 | 5 | -1.666667 | 5.3333333 | -1.777778 |
| 3.5 | 6.902E-09 | -2.96E-08 | 1.352E-07 | -6.57E-07 | 4.2857143 | -1.22449 | 4.5714286 | -1.306122 |
| 4 | 9.313E-10 | -3.49E-09 | 1.397E-08 | -5.94E-08 | 3.75 | -0.9375 | 4 | -1 |

Table 5: Properties of the first four derivatives of the four utility functions used in this chapter at seven values in the range [1,4] with the coefficient of relative risk aversion equal to 15 at the point of next period's expected consumption in each case.

are mimicking. These figures can be directly compared with four of the six graphs presented by Weil (1992a) on pages 786–7. Notice first that the “to $U^{(3)}$ ” utility has a predicted equity premium very close to that of the representative agent model for all values of Gamma. So, despite the fact that this utility function is globally Standard Risk Averse there is little change in the risk aversion of the agent through the introduction of unemployment risk. This is the initial indication that the “very large factor” underprediction of the representative agent model is dependent on stronger assumptions than just DAP and DARA.

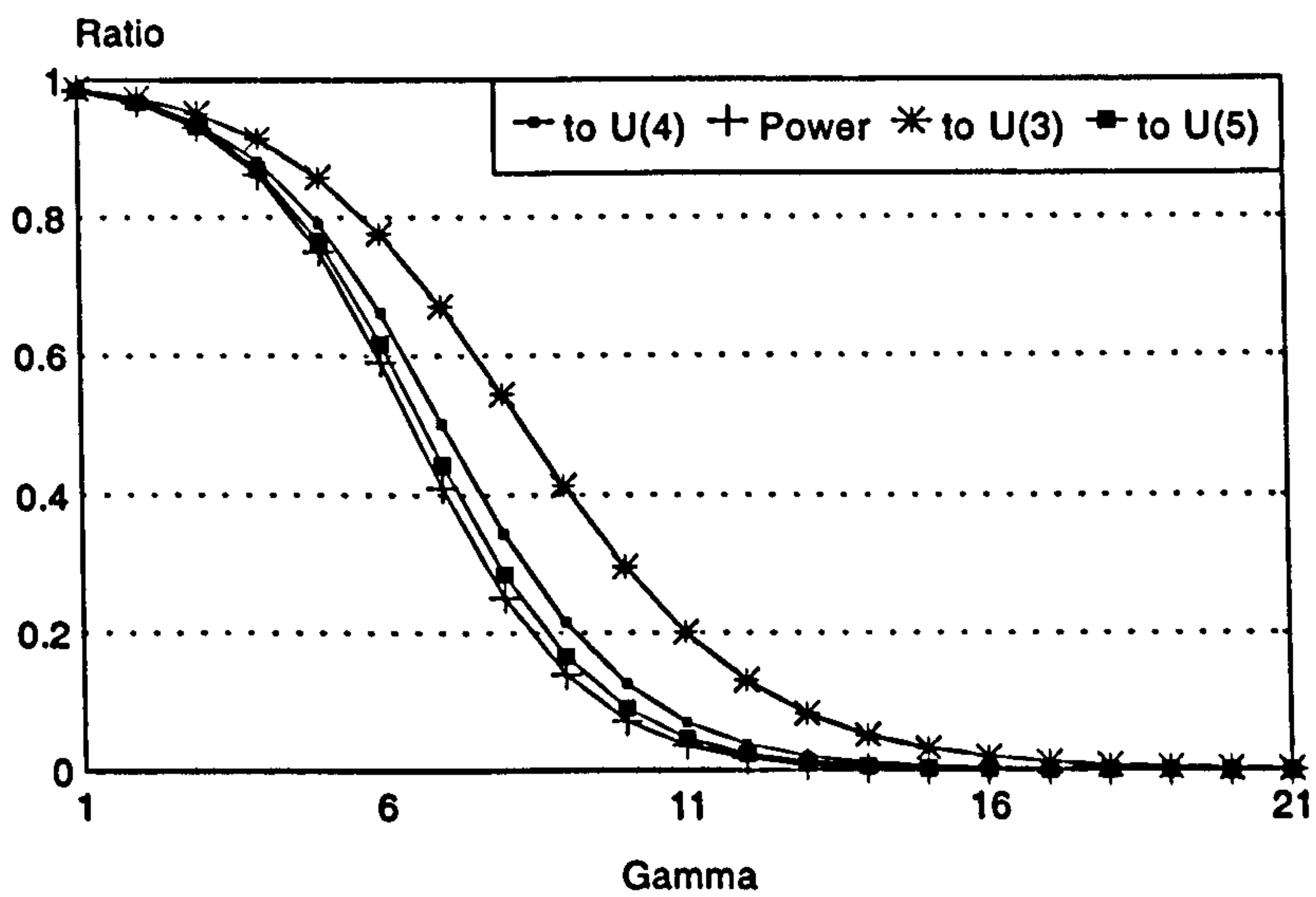
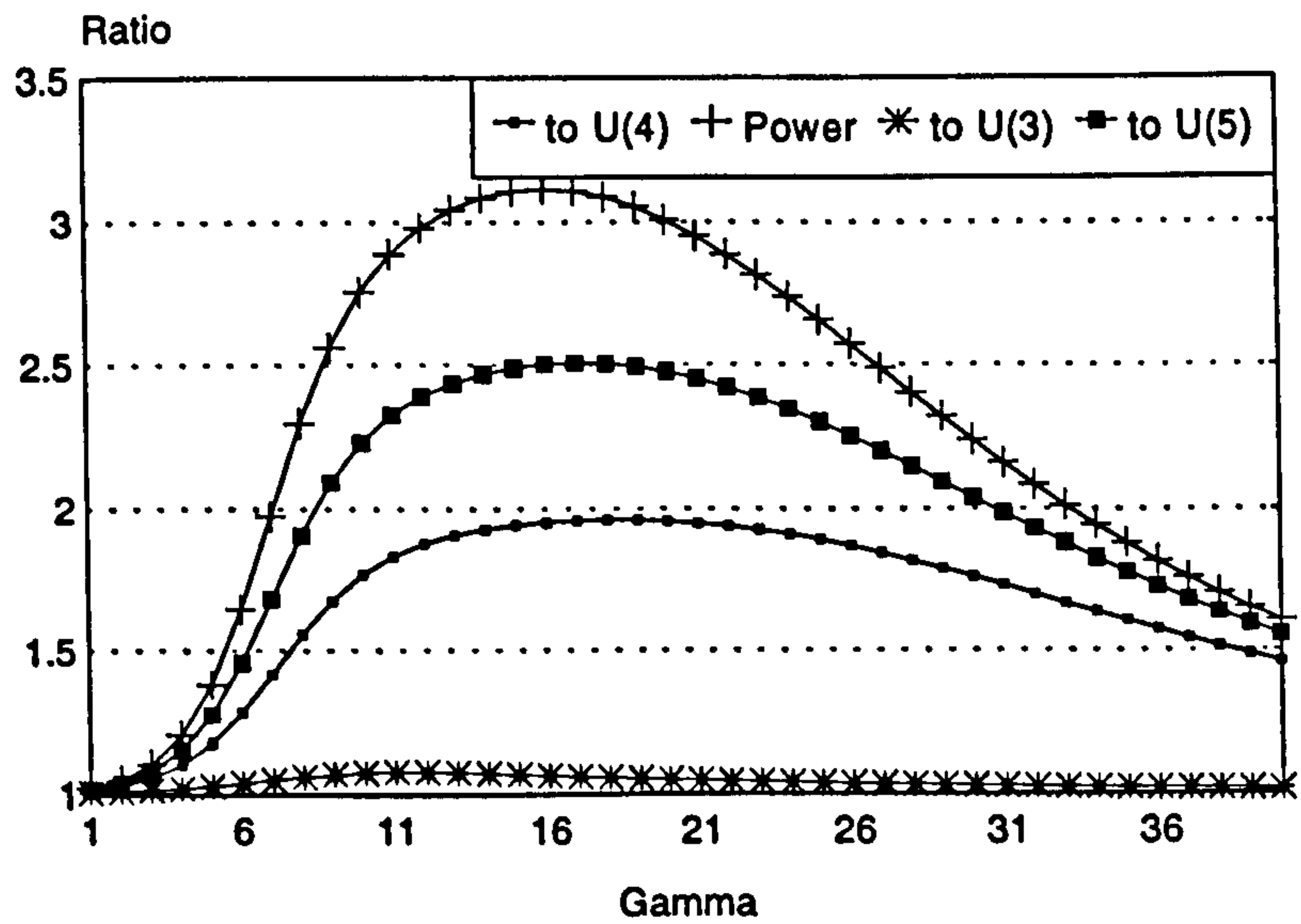


Figure 8: These graphs replicate and adjust two of the three graphs on page 786 of Weil (1992a). Unemployed income is one third the employed income. The top (bottom) graph is the ratios of the predicted equity premium (riskfree rate) in the presence of unemployment risk to the predicted equity premium (riskfree rate) in the absence of unemployment risk.

The riskfree rate is significantly altered for this utility function by the introduction of unemployment risk as the demand for savings is driven by the precautionary savings motive: that is by prudence \mathcal{P} and not \mathcal{A} . Consider next “to $U^{(4)}$ ” and “to $U^{(5)}$ ”. For these, the ratio of predicted equity premia in the presence and absence of unemployment are similar to the same ratio for power utility for high levels of relative risk aversion. However for intermediate values of γ , the ratios are significantly lower. For example, in the case when unemployment income equals 1, the highest ratio is around 2 and 2.5 respectively compared with a ratio of over 3 for power utility. When unemployment income equals 0, the highest ratio is around 2.5 and 4 respectively compared with a ratio of over 9 for power utility. Therefore there is a significant change in the predicted equity premium compared to the representative agent model under the same utility function but not to the same extent as with power utility. Given that these two utility functions are identical to power utility locally to the fourth (fifth) derivative, this implies that the magnitude of the equity premium is significantly affected by very high order derivatives for negatively skewed gambles such as Weil’s. The importance of these higher order derivatives in the quantitative (as opposed to qualitative) assessment of the equity premium are of great importance.

21 Conclusion

By using certainty equivalents and concentrating on small consumption gambles the author believes that he has contributed to existing reviews by intro-

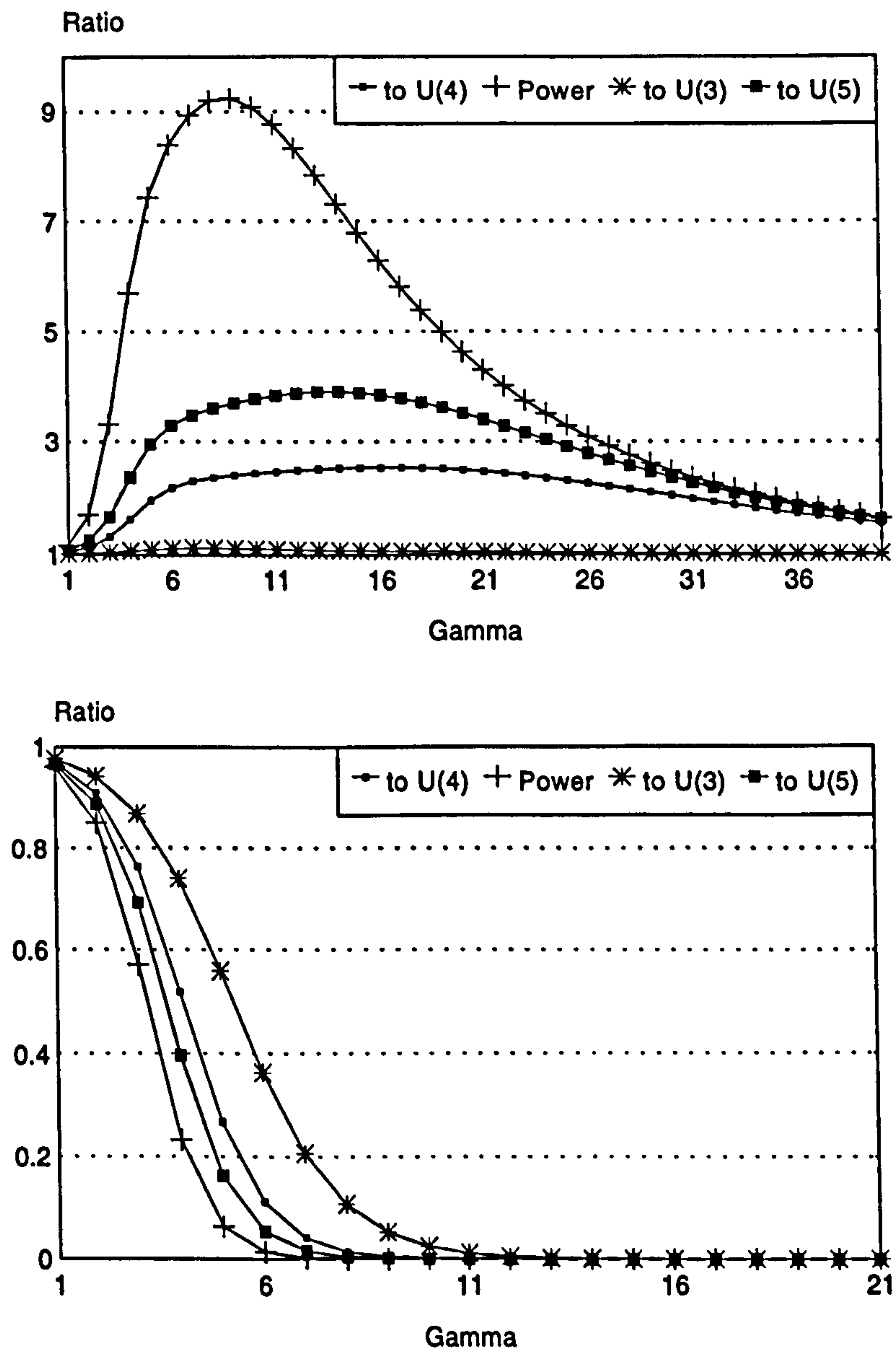


Figure 9: These graphs replicate and adjust two of the three graphs on page 787 of Weil (1992a). Unemployed income is zero. The top (bottom) graph is the ratios of the predicted equity premium (riskfree rate) in the presence of unemployment risk to the predicted equity premium (riskfree rate) in the absence of unemployment risk.

ducing a more integrated framework for understanding the different forms of proper risk aversion. As a consequence of this approach, a new and highly general form of proper risk aversion — Basic Risk Aversion — follows naturally. If the utility function is restricted to have positive third derivative and the investor is Basic Risk Averse then the investor is certainly proper risk averse under all existing forms of proper risk aversion. Despite this, the necessary and sufficient conditions for local Core Basic Risk Aversion are no more restrictive than for local Standard Risk Aversion. Therefore DAP and DARA will ensure that a greater class of small consumption gambles than has previously been recognised will increase the aversion of an investor to an independent marketable gamble.

Weil (1992a) makes an important contribution to the Financial Economics literature by showing how the introduction of independent uninsurable income risk can make a dramatic impact on equilibrium asset pricing when utility is assumed to take power form. He also provides robust theoretical justification for this phenomena through the properties of decreasing absolute risk aversion (DARA) and decreasing absolute prudence (DAP) which ties his paper in with an established literature on proper risk aversion. This chapter contributes to this debate by comparing the theoretical literature that predicts an increase in the equity premium with the example of Weil that gives the magnitude of this effect for one specific example. The magnitude of the effect will depend on the ratio of the expected value of the second derivative to the expected value of the first derivative. This can be

calculated in one of two ways. Either the derivatives are estimated explicitly over the consumption domain or the utility function is assumed to follow a given equation. If, as with Weil's example, it is assumed that the utility function can be described by an equation with fixed parameters, then it is must be decided to what precision the functional form of the equation must be estimated. From the literature on proper risk aversion, through its emphasis on the first four derivatives of the utility function, the reader might infer that concentrating on these derivatives alone might provide a reasonable approximation for the magnitude of the effect of proper risk aversion. This chapter has shown that this inference is not supported for Weil's example. By creating alternate utility functions that are well defined and well behaved (in terms of monotonicity, concavity, DAP and DARA) across the domain and identical locally to power utility to the third, fourth or fifth derivative at the point of next period's expected consumption it is shown that very different equity premium can be predicted. So if we are interested in the quantitative (as opposed to qualitative) effects of the introduction of exogenous uninsurable background risk in an example like Weil's and are going to use a utility function described by a simple equation then it is necessary to get introspective knowledge of high order derivatives of this equation.

Part V

Uninsurable risk and optimal dividend policy

Uninsurable risk and optimal dividend policy⁸⁹

Abstract

In this chapter, the role that risk to individual income plays in determining the attractiveness of a claim that has stochastic financial obligations (“investments”) as well as payoffs is examined. It is shown that investors prefer economies where the uncertainty about future investments is lowest in states with the greatest risk to personal capital. It is argued that this is consistent with an optimal financing policy where dividends are smoothed and right issues are a bull market phenomenon. It is also contested that observed corporate finance behaviour is consistent with the incomplete market models that lie at the heart of this thesis.

⁸⁹A version of this chapter was presented to the Doctoral Colloquium, European Finance Association, Milan, August 1995 and the Financial Options Research Center, University of Warwick, November 1995

22 Introduction

In the previous chapters of this thesis, the role that idiosyncratic endowment shocks might play in resolving the puzzles of Mehra & Prescott has been discussed in some detail. In chapter 6 a series of empirical tests are run that examine whether the change in asset returns with variations in the risk to personal capital is consistent with these models. This chapter considers an entirely separate market anomaly using the technology of incomplete market theory that has been discussed above. Rather than examining the Mehra & Prescott puzzles, the dividend controversy discussed most notably by Black (1976) is considered in an incomplete market framework. This is, to the author's knowledge, the first attempt to use incomplete market theory to address problems in corporate finance. Strictly speaking, it is an *aggregate* optimal dividend policy that will be constructed⁹⁰. It will be argued that the observed corporate finance policies of individual firms that comprises of dividend smoothing and concentrating rights issues in bull markets is consistent with the optimal economy-wide practice that emerges from the models in this chapter.

This chapter follows in the spirit of the Mankiw (1986) model described in chapter 3. At $t = 0$ all investors are *ex-ante* homogeneous. At future

⁹⁰While this is, perhaps, somewhat unusual, there are precedents: *cf.* Miller (1977) on optimal capital structure with personal and corporate taxes. This type of macro solution to the dividend puzzle is also hinted at by Marsh and Merton (1987): "For example, in a purely demand-driven model for dividends, the demand for dividends is not firm specific because investors only care about the dividend-capital gain mix at the portfolio level ... Thus equilibrium aggregate dividends may be determinate, but which firms service this demand and the quantity each chooses to supply may not" (pp.4-5).

time periods there are two states of the world, an upstate and a downstate. In the upstate there is no risk to personal capital. In the downstate there is the possibility of unemployment. Previous chapters show that the risk to individual income combined with risk to dividends increases the risk aversion of investors to the market and so the equity premium puzzle can, at least theoretically, be resolved in this way. This chapter is *not* primarily concerned with dividend / profit risk. Instead this chapter focuses on the effects of combining endowment shocks with *investment* uncertainty. That is, the main “worry” for investors within this economy does not concern the return on capital but is, instead, the possibility that aggregate investment will not always be at the optimal level⁹¹. The less information investors have about future investment plans, the more concerned they are about it being suboptimal. It is shown that this effect is amplified in states where there is also a risk to personal capital.

The corporate finance implications of the findings in this chapter are as follows. There are two effects. For DARA utility functions, investors prefer lower investment in the low state to the high state as there is less endowment to fund investment in the former case. So, *positive* shocks to investment should be concentrated in high states. As higher than expected investment will coincide with lower than expected net payouts (dividends - rights issues) to investors within this type of model this means that rights issues should

⁹¹Investment is “optimal” in this context if managers cannot increase the expected utility of consumption of at least one investor and not reduce the expected utility of consumption of any other investors by changing investment from this level.

be mainly bull market phenomena and dividends should not be cut in bear markets. This argument about the optimal *level* of investment in different states is, though, reasonably straightforward and does not require the intricate simulations contained within this chapter. Here, in an environment where expected investment is the same in both states, the role of mean zero investment shocks is examined. These are undesirable in low endowment states as this coincides with a period when there is also high risk to personal capital. The combination of these two risks is examined below. So, if mean zero investment shocks are less desirable in low states than high states, and if the optimal level of investment is greater in the latter case than the former, then it is argued that a corporate finance policy of smooth dividends combined with rights issues concentrated in high endowment states is consistent with the model developed here. Paying dividends and having a rights issue in the same year may also be justifiable within this type of model. Essentially dividends are smoothed because the company is saying that “this is the minimum cash flow that you will receive from your investment at times of low endowment. We will not have a rights issue at these times to reclaim this cash from you.” Rights issues are concentrated in bull markets as mean zero investment uncertainty and positive investment shocks are most easily absorbed in these states.

Within the economy to be described there are two types of individual: managers and investors. This is not, though, an agency model. The simulations are constructed in such a way that managers are rewarded by aiding

investors. The simulations show that, by working in their own interests, managers also maximise the expected utility of investors. That is, purely benevolent managers and purely selfish managers would make the same decisions in all the cases considered. Therefore we do not need to worry about the usual manager / investor “games” that sometimes appear in the corporate finance literature.

This chapter is an innovative application of incomplete market theory. In order to fully motivate the work to follow, there now follows a long introductory discussion. First, the two-period version of the model is described in depth. Then, the key assumptions that drive the model are drawn out and discussed. Third, the work is placed in the context of other theories of optimal dividend policy. Finally, the testable implications of the model are drawn out and existing empirical evidence is reviewed in the light of this new theory.

22.1 The economy

This chapter works in both a multiperiod and two period world. A brief description of the two period model is now given. The multiperiod model is a natural extension of this:

- There is one firm in the economy (the “market portfolio”) which is all equity financed. There is no real investment either before $t = 1$ or after $t = 2$.

- At $t = 0$ managers tell investors that they have a real investment opportunity that will give constant returns to scale of r which may, or may not, be known with certainty. They announce that they will have a rights issue of approximately (say) £1 per share at $t = 1$. There are no other investment opportunities available.
- The assumption that drives the simulations is that the true size of the rights issue at $t = 1$ may not be revealed at this stage. That is, at time 0, investors are uncertain about time 1 investment per claim.
- At $t = 0$ all parties know that in the next period the economy will go into either a high state h or low state l . At $t = 1$ the true size of the rights issue will be $£1 + e_x$ ($x \in \{h, l\}$) where e_x is a mean zero random draw whose variance is state dependent. The managers of the firm have control over the volatility of e_x but not its expectation. That is managers can choose to “shock” investors over the size of the rights issue but cannot choose a positive or negative shock. Within this model, the constraint is that investors will be shocked if the economy turns out to be in one state but not if it is in the other. That is $e_h = 0$ or $e_l = 0$ with certainty and e_x will have non-zero variance in the other state. Managers reveal to investors at $t = 0$ which state contains the investment uncertainty and which state does not.
- On the basis of this information the investors choose the number of shares k that they wish to take up. By having complete control over

k , investors decide on the size of real investment at $t = 1$ ⁹². As there is *ex-ante* homogeneity within this economy, all investors choose the same k . At $t = 0$ this involves no investment at this stage but there is commitment for the future.

- At $t = 1$, each investor receives exogenous endowment. If the state is high then all investors will receive the same income. If the state is low then some will continue to receive this endowment but some will get a much lower level of income (“unemployment benefit”). *Ex-ante* all investors have the same probability of becoming unemployed should the state be low at $t = 1$. Whatever their income level, each investor must then pay $k(\pounds 1 + e_x)$ to the firm or else sell some securities on to another investor who will fulfil the obligation on this portion of the portfolio⁹³.
- At $t = 2$ each investor receives the same endowment as at $t = 1$. That is, unemployment is persistent. The firm returns a dividend $(\pounds 1 + e_x)(1 + r)$ per claim and the economy terminates.
- The point of contention is “In which state should investors’ receive the investment shock?” It is argued in this chapter that the combination of potential low income combined with an investment shock will make investors more wary of the company if investment is uncertain in the state with high risk to personal capital.

⁹²An equivalent way of modelling this would be to allocate each investor with one share and then let investors choose the expected size of rights issue per claim.

⁹³In the multiperiod world there is no opportunity for this type of trade.

As this is the first model of its kind in the literature, it may be helpful at this point to demonstrate the effects with an example. In this simplified case, trading is prohibited. At $t = 0$ the one firm modelled announces that it has a real investment opportunity offering fixed return to scale of 5%. Managers tell investors that the expected size of the rights issue will be 0.05 units of the consumption good per claim. In one state the size of the rights issue will certainly be 0.05, while in the other state investment will either be 0 or 0.1 with equal probability. Investors are told which state contains the investment shocks and they then choose how many shares they wish to hold (k). At $t = 1$ the economy goes into either a high or low state. In the high state all receive income of 1 unit of the consumption good. In the low state 90% of the population continue to receive this endowment while 10% become unemployed and receive only 0.4 units. Each individual is given the same endowment at $t = 2$ as $t = 1$. If all investors have power utility with $\beta = 1, \gamma = 3$ ($U(c) = -0.5c^{-2}$) then the optimal number of claims is $k = 0.0518$ if $e_l \in \{-0.05, +0.05\}$ and $e_h = 0$ and $k = 0.0807$ if $e_l = 0$ and the investment uncertainty is in the high state. The respective expected utility of future consumption is $-1.26239, -1.26233$. Investors prefer investment uncertainty in the high state as their expected utility of future consumption is higher in this case (if only by a small amount). Managers also prefer investment shocks in this state as the expected value of their firm is also higher in this case. This is what is meant by there being no agency conflicts in this model. Selfish and benevolent managers take the same action — to

concentrate aggregate investment shocks into states of low risk to personal capital.

22.2 The key assumptions

In this section, the key assumptions that drive the model are discussed and it is argued that these assumptions are reasonable. The main driving assumption is that investors have uncertainty about the future levels of real investment undertaken by firms in a well diversified portfolio and, also, they cannot always rely on the managers of the firm to invest optimally on their behalf. This is rather like Jensen's (1986) idea that managers have a systematic tendency to overinvest. In this chapter, though, deviations from the original investment plan have zero mean and so this model does not rely on agency theory⁹⁴. There are two parts to this assumption: (i) aggregate investment is uncertain and (ii) aggregate investment is sometimes suboptimal. These are dealt with in turn. Consider the top of figure 10 which shows fixed investment by manufacturing companies in the UK from 1955–96. This time series appears unpredictable and volatile. Certainly, at the microeconomic level, the investment plans of individual firms are unknown as “...one of the “inside” variables that a firm cannot readily communicate without moral hazard is the level of new investment” (Bhattacharya (1979) p. 261) as “... the firm may not be able to announce its investment plans because of competitive reasons...” (Ang (1987) p. 45). The top graph in figure 10 suggests that there is a systematic component in the difference between

⁹⁴Although it would be possible to construct an agency story around these simulations.

actual and expected investment across firms (as might be predicted by, say, a fads theory of investment) leading to aggregate uncertainty.

It may be that the unpredictable pattern of the top graph of figure 10 just reflects the changing optimal level of real investment. Managers amend their investment plans at each point in time to keep the level of investment optimal. While it is not possible to reject the hypothesis, it does not seem economically reasonable. Optimal investment rates change instantaneously as investors receive endowments and change their expectations of future endowments. Real investment, on the other hand, has a large degree of irreversibility — this point lies at the heart of the real options literature: see, for example, Dixit and Pindyck (1994) — and new projects take considerable planning. Indeed, looking again at the top figure of 10, it seems that real investment does not fall (rise) as soon as discount rates rise (fall). This supports the anecdotal evidence that there is a significant lag between changes in interest rate and real economic activity. Further, it would be possible to construct an agency story around these simulations where investment is suboptimal because managers are too lazy to cancel negative NPV projects or find new positive NPV projects in which to invest. Therefore to assume that future real investment may be suboptimal, which is the key driver of this chapter, appears to be an economically plausible assumption.

If each individual managed their own portfolio of real projects in addition to equity in the firm then, by adjusting their own real portfolio, optimal investment levels could be maintained. This is assumed away by making

the equity in the firm the only investment opportunity: “...the vision of an economy dominated by large corporations that are owned by portfolio investors has in large measure been realised, at least by the most advanced capitalist countries, so that this assumption may well be considered the most innocuous of the five stated earlier⁹⁵” (Gordon (1996) p.14).

Finally, there is an assumption that the expected level of investment is state independent: investment shocks have zero mean in both cases. It is reasonable, though, to suppose that investors would prefer more investment in the high state than the low state as discussed above. For an investor with power utility who is offered certain return r on investment i over the interval $[1, 2]$ and has known endowment of y at both times, the optimal value i is:

$$i = y \left[\frac{1 - (1 + r)^{-1/\gamma}}{1 + (1 + r)^{1-(1/\gamma)}} \right]$$

So, the higher the endowment the higher the desired level of investment⁹⁶. This means that the employed will want to invest more than the unemployed. Therefore, for power utility the optimal strategy for the manager is to change the expectations of e_x, e_l to make the former positive and the latter negative.

However:

⁹⁵The five propositions that underlie the neoclassical theory of finance and investment.

⁹⁶Although this is not true for exponential utility function $U(c) = -\text{Exp}(-ac)$ when the optimal $i = \ln(1 + r)/a(2 + r)$, which is independent of y . If i is small compared to y then the optimal $i \approx r/A[1 + (1 + r)^2]$ so this result is a feature of DARA. The effect is amplified if the state can change between times 1 and 2. This is because, in the high state, investors will wish to precautionary save against the prospect of becoming unemployed next period. In the low state, while this effect still occurs for those who remain employed, it is offset by the unemployed who will be prepared to save less in the hope of becoming reemployed next period.

- While the expected level of real investment should be lower in low states than high state, the difference is not very great. As can be seen in the previous offset equation, the optimal level i is linear in y . If the risk of unemployment is 5% and unemployment benefit is 40% of employed income, optimal investment in the low state will only be 3% lower than in the high state⁹⁷. Further, as for most of the population, endowment is the same in both states, the optimal level of investment per share is state independent for most investors. Therefore if managers listen only to the “majority view” then the level of investment will be state independent. Also the argument given above about the time delay between planning a project and undertaking a project means that companies will not be able to switch from “big” projects to “small” projects (or vice-versa) at the time the state reveals itself.
- As mentioned above, the implications of this chapter are *strengthened* by having the mean value of e_x state dependent. The mean of e_l is optimally lower than the mean of e_x . So times of *positive* investment shock should be in the high states. Therefore, on this expectations argument, rights issues and dividend cut, which will be used to fund high investment, should be concentrated in high states. Section 24.2 corroborates this. In this case, the probability of a rights issue is made state dependent. It is shown in this case that the results are amplified. The other simulations extend this result to show that *even if* the

⁹⁷Indeed, if expected endowment is the same in both states then the allocation of income shocks does not alter optimal aggregate investment as i is linear in y .

investment shock has zero expectation then it is still more favourably received in a high state than a low state.

22.3 The theory of dividend policy

This subsection aims, very briefly, to present the model just described in the context of existing theories of optimal dividend policy. This subsection does *not* attempt to be exhaustive. While the author feels that chapters 2 & 3 make a substantive contribution to the literature by providing more comprehensive coverage of the Mehra & Prescott puzzles than any previous study, an equivalent chapter on dividend policy would not make such a contribution. This is because major overviews of the dividend policy literature already exist. See, in particular, Ang (1987) but there are several from which to choose⁹⁸. The reader is referred to these sources for more detail.

The closest model to the one developed here is the overinvestment theory of Jensen (1986). In his economy, there is a systematic tendency for the firm to overinvest as this is in the interest of managers. Even though in the simulations in this chapter investment uncertainty has zero expectation, it is still the danger of overinvestment that is driving the results. Essentially the threat of overinvestment at a time when endowment is low is the dominating factor here. This is because the company will either have to pay a lower dividend than expected or have a rights issue at a time when some individuals have lower income than usual. Because this threat is so strong, it is not

⁹⁸See, for example, Miller (1986), a set of articles in the *Journal of Economic Perspectives* (1988), Weston (1989) and Gordon (1989), amongst others.

necessary to have a systematic risk of overinvestment. So it is the possibility of suboptimal investment that gives the results of this chapter rather than systematic overinvestment. This means that it is not necessary to appeal to agency problems to motivate the economy. Indeed, as stressed above, the economy is constructed in such a way that agency conflicts are removed. It is contested that companies pay dividends in each year to “signal” to investors the level of payouts that they will receive in future low states. Rights issues are added in high states to fund investment commitments. So dividends in high states are conveying information about net cash payouts in low states. How does this compare with a standard signalling model? “If one accepts that dividends convey new information, it is not certain which of the following types of “new” information is sent. It could be a forecast of future earnings ... a more precise estimate of current earnings ... an indication of the permanence of past earnings” (Ang (1987) pp. 39–40). In this chapter, the new information conveyed is about future *dividends* and not earnings. Further, a number of problems with the more traditional dividend signalling models are overcome. Again, Ang highlights three difficulties with signalling (i) why signal with dividends? (ii) who pays the price for false signals? (iii) signalling can become self-fulfilling — managers won’t cut dividends because they are afraid it will be seen as a signal so investors view dividend cuts as a bad signal. In this model, dividends are being used to convey information about future dividends, which seems natural, although there may still be less expensive ways of conveying this information. False signals are punished by

sharply falling share prices following dividend cuts, which has been widely reported. In this economy managers gain benefit from maximising the value of their firm. In this model investors don't like dividend cuts because of the cash-flow implications and the self-fulfilment of the usual signalling model is therefore removed. So this model is not heavily based on the traditional dividend signalling models of Bhattacharya (1979), Hakansson (1982), Miller and Rock (1985), John and Williams (1985) and Bar-Yosef and Huffman (1986).

Finally, this chapter has a model with endogenous investment and financing decisions. This brings to mind the literature started by Myers and Majluf (1984). In their model, companies should retain profits rather than pay dividends if there are potential future positive NPV investment opportunities as issuing new shares can result in a redistribution of wealth from existing shareholders to new shareholders. Again, this chapter is clearly not based on this type of model. So, it is argued that, while the economy described in this chapter has echoes of other theories of optimal dividend policy, this is a “new” approach. Indeed, this chapter is based more heavily on the work on incomplete market theory described and developed in the other parts of this thesis than existing studies on corporate finance tactics.

22.4 The empirical evidence

Finally, some existing empirical studies on the impact of dividend policy are interpreted in the light of the model described above and simulated below.

As with the whole of this section, there is no specific attempt to formally

reject the model. Instead, the aim is to motive the reader into believing that this is a model that merits detailed future empirical investigation.

The main testable implication of this model is that dividend cuts and rights issues will be most unfavourably received at times when risk to personal capital is high. This is partly because positive investment shocks are best received at times when endowment is high. It is also because the simulations below will show that, even if the investment shock has mean zero expectation, then it is still better absorbed in the high state. So, if an event study of abnormal performance after a rights issue or dividend cut were to be run then it should be expect that the residuals would be at their most negative at times of high individual risk. Unfortunately, due to the procyclical nature of the volatility in individual income, this makes the model quite difficult to distinguish from other business-cycle based models of corporate finance policy. Despite the many studies on market responses to dividend changes there appears to be no work on how these reactions change over time. The bottom graph in figure 10, based on evidence given in Sant and Cowan (1994), shows the number of dividend omissions in the US over the interval 1963–86 and the level of unemployment at the time (see the next chapter for more details on this time series). This shows that the number of dividend omissions often rise at times of high unemployment (1974, 1982). This is not direct evidence against the model presented here, however. As high unemployment often coincides with economic recession it might be reasonable to expect more refinancing by firms at times of high unemployment. With regard to

the timing of rights issues, Choe, Masulis and Nanda (1993) show that rights issues are more prevalent in bull markets than bear markets, although the difference is not great. In real (1983) terms, the annual average value of rights issues in the US was \$12.4bn in upmarkets and \$10.1bn in down markets over the interval 1971–91. Further, the market's response to rights issues is less negative in times of economic boom than economic recession. Unfortunately these results are not only consistent with the simulations presented above — see Choe et al. (1993) for alternatives. Therefore, while the implications of these results are ambiguous for the model presented here, they are at least consistent with our economy.

There has been a group of papers (for example, Lang and Litzenberger (1989), Yoon and Starks (1995), Denis, Denis and Sarin (1994)) that have examined whether the empirical evidence is more supportive of Jensen's over-investment model or a cash-flow-signalling model. On balance, the evidence does not support Jensen's model (although Lang and Litzenberger (1989) does argue in its favour). This is because both Yoon and Starks (1995) and Denis et al. (1994) find evidence of investment increasing after dividend increases⁹⁹. Therefore an increased dividend is not a sign of lower overinvestment. This, though, does not conflict with the model presented here as an increase in dividend is just increasing the commitment of the company to higher net payouts in states with high risk to personal endowment. It says

⁹⁹Pruitt and Gitman (1991) find in a survey of 114 firms that "... the dividend decision was found to be made independently of the firm's investment and financing decision" (p.428)".

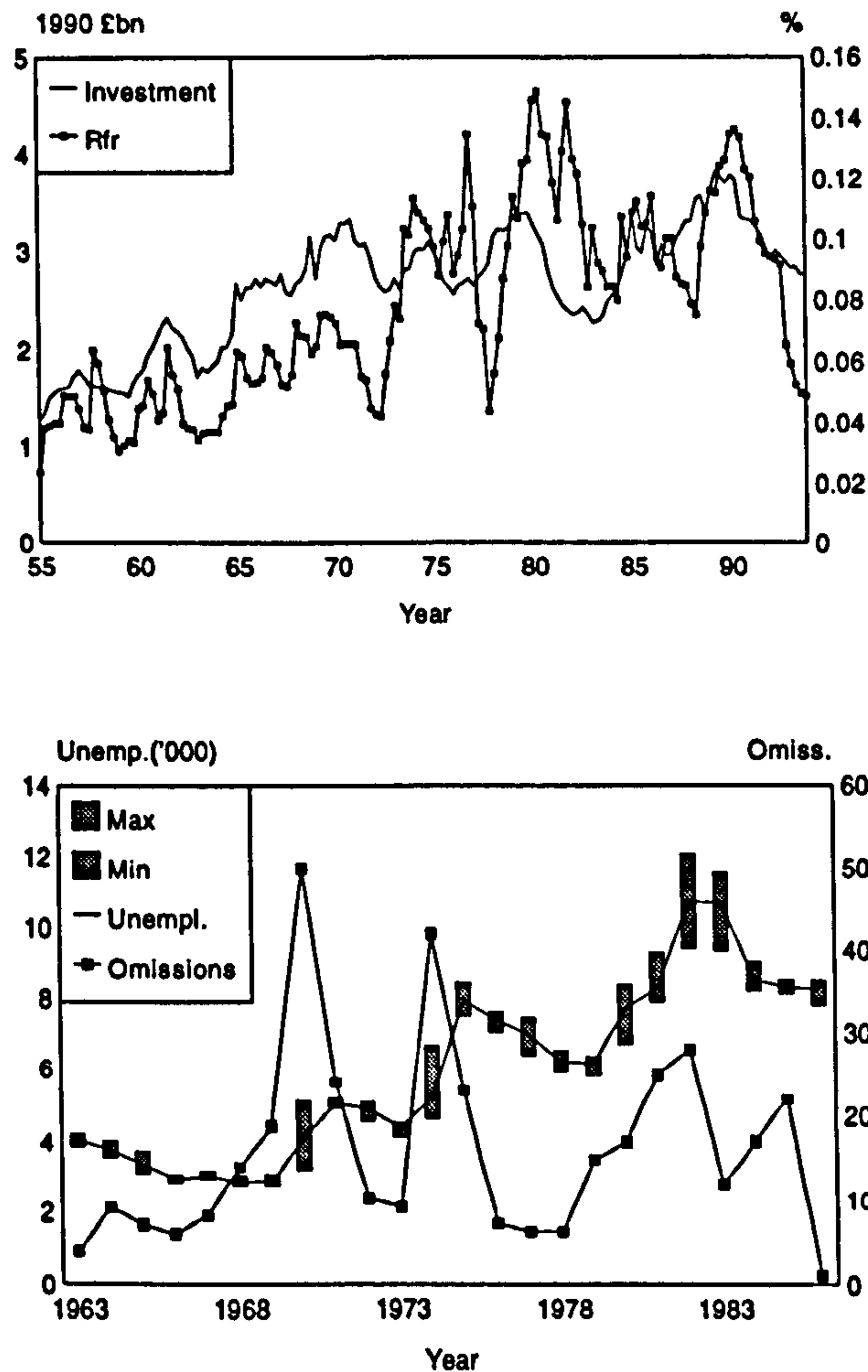


Figure 10: Top graph: fixed investment by manufacturing companies including leasing in 1990 in the UK (left hand axis) and the nominal riskfree rate (right hand axis) 1995Q1 – 1993Q4. Source: Office of National Statistics and the LSPD. Bottom graph: the number of dividend omissions against unemployment for the US 1963–86. Data on dividend omissions is described in Sant and Cowan (1994) and unemployment data is discussed in the next chapter. Dividend omissions should be read against the right hand axis. The max-min bars, which should be read against the left hand axis, give the range of monthly unemployment over the year. The Unempl. line gives average monthly unemployment in that year.

nothing about the future level of investment. Also, it is the smoothness of aggregate dividends in states with a high risk to personal income that is important rather than the cross-sectional sourcing of this dividend. Therefore this model has the advantage of being consistent with the empirical findings of Smith and Watts (1992) and Gaver and Gaver (1993) who find that dividend yield is inversely related to investment opportunities.

Finally it is clear why the model presented here would suggest a smooth dividend policy as famously found by Lintner (1956). The dividend of a company is giving a commitment to investors as to the minimum amount of net cash flow at times of economic recession. This commitment can be increased steadily as the firm grows but may not be cut. So, in conclusion, existing studies are consistent with the results of the simulations in sections 23 and 24.

The chapter proceeds as follows. In section 23 numerical methods are used to describe a two period economy. The reason for concentrating in a two period economy is that, at times of high unemployment, the unemployed might still be able to make “home-made dividends” through selling their assets. By simplifying to a two-period model the economy is sufficiently simple so that numerical methods can determine the price and quantities of trading in the market. Within a multiperiod economy calculating the optimal trading rules is extremely complex as the number of potential paths through the economy are so many. The multiperiod economy is left to section 24 and there will be no trading in this case. The *contribution* of this chapter is

clear. This is the first application of a model in the style of Mankiw (1986) to issues in corporate finance. While additional empirical tests are needed, the existing evidence appears to be unable to reject the model. Further, the intuition behind this model is highly appealing. That companies pay dividends at times of low profitability because investors require the cash flow to smooth their endowment shocks has, to this author, more elegance than standard signalling / taxation models. It is likely that a similar technique could be applied to issues of optimal (aggregate) capital structure. It is not contended that all issues are resolved, but this does appear to be a fruitful route for further investigation.

23 A model with trading

23.1 Describing the economy

In this chapter both two period and multiperiod models are examined. In the former models there is trading so that individual consumption can be partially smoothed by buying and selling assets. The long term implications of a particular investment strategy is better examined in the latter class of models even though it will not be possible to construct trading rules in this case.

This section deals with two-period models with trading. This model was described in some detail in the introduction, so here the aim is to convert this into notation and formally define the problem. As before, there is one firm in the economy (the “market portfolio”) which is all equity financed.

There is no real investment either before $t = 1$ or after $t = 2$. At $t = 0$ managers tell investors that they have a real investment opportunity that will give constant returns to scale of r which may, or may not, be known with certainty.

- At $t = 0$ all parties know that in the next period the economy will go into either a high state h or low state l with equal probability. They know that at $t = 1$ there will be a right issue of size $\tilde{i} = i + e_x$ ($x \in \{h, l\}$) per claim where e_x is a mean zero random draw whose variance is state dependent. Managers can choose to have $e_l = 0$ or $e_h = 0$ with certainty (but not both). Investors know: (i) i , (ii) whether e_l or $e_h = 0$ and (iii) the distribution of investment uncertainty in the other state (which does not depend on whether e_l or $e_h = 0$). Because each state is equally likely to occur, the overall investment uncertainty is not decided by whether e_l or $e_h = 0$.
- On the basis of this information the investors choose the number of shares k that they wish to take up. As there is *ex-ante* homogeneity within this economy, all investors choose the same k . Use k_h^* (k_l^*) to denote the number of shares investors will choose if $e_h = 0$ ($e_l = 0$). Managers have sufficient information to be able to calculate these two values of k at $t = 0$.
- At $t = 1$ the state of the world is determined. If the state is high then all investors receive income y_h . If the state is low, then a proportion

$1 - s$ receive endowment y_h , the remainder endowment $y_u \ll y_h$. *Ex-ante* all investors have the same probability s of becoming unemployed should the state be low at $t = 1$. The true size of the rights issue \tilde{i} is then revealed.

- In the low state, a financial market then operates to enable the unemployed to trade shares with the employed. Such a market is redundant in the high state as all investors are homogeneous. The unemployed will be able to sell l of their k shares to the employed at a price p immediately prior to the firm claiming \tilde{i} . Define $S := s/(1 - s)$ so that then Sl is the number of claims purchased per capita by the employed from the unemployed. A negative value for l merely indicates that the unemployed investor is a buyer rather than a seller. It is clear that, in this case that the consumption in the high state (c_{h1}), low state, employed (c_{e1}) and low state, unemployed (c_{u1}) at $t = 1$ is given by:

$$\begin{aligned} c_{h1} &= y_h - k\tilde{i} \\ c_{e1} &= y_h - Slp - (k + Sl)\tilde{i} \\ c_{u1} &= y_u + pl - (k - l)\tilde{i} \end{aligned} \tag{49}$$

- At $t = 2$ all investors continue to receive the same endowment as they received at $t = 1$. The firm pays a dividend $\tilde{d} := (1 + r)\tilde{i}$ per claim and the economy terminates. The consumption in the high state (c_{h2}), low state, employed (c_{e2}) and low state, unemployed (c_{u2}) at $t = 2$ is given by:

$$\begin{aligned}
c_{h2} &= y_h + k\tilde{d} \\
c_{e2} &= y_h + (k + Sl)\tilde{d} \\
c_{u2} &= y_u + (k - l)\tilde{d}
\end{aligned} \tag{50}$$

Two issues remain unresolved. How do the investors choose k and how do managers decide whether to let $e_h = 0$ or $e_l = 0$. The choice of k comes from investors maximising their (time zero) expectation of utility of consumption $E_0[u(c_1, c_2)]$. It is assumed throughout that investor preferences are additively time separable with no time preference so that $E_0[u(c_1, c_2)] = E_0[U(c_1)] + E_0[U(c_2)]$. Formally:

$$\begin{aligned}
k_h^* &:= \max_{k|e_h=0} E_{i,r}[U(c_{h1}) + U(c_{h2}) + s[U(c_{u1}) + U(c_{u2})] + \\
&\quad (1 - s)[U(c_{e1}) + U(c_{e2})]] \\
k_l^* &:= \max_{k|e_l=0} E_{i,r}[U(c_{h1}) + U(c_{h2}) + s[U(c_{u1}) + U(c_{u2})] + \\
&\quad (1 - s)[U(c_{e1}) + U(c_{e2})]]
\end{aligned} \tag{51}$$

The baseline case will be for utility to be of power form, although exponential and quadratic utility are also considered in some of the models.

Finally it is necessary to decide how managers will choose whether e_l or $e_h = 0$. There are two “obvious” alternatives. First the manager could choose to act on behalf of investors and maximise investors’ expected (time zero) utility of consumption. That is $e_l = 0$ would be set equal to zero if the expected utility with investment uncertainty in the high state and k_l^* shares issued is greater than with investment uncertainty in the low state and k_h^* shares issued. Alternatively managers gain benefit from maximising the size of the expected value of their firm at $t = 1$. This is a standard agency-style concept as encapsulated by Jensen’s model. As the expected level of

investment per claim is the same whether e_l or $e_h = 0$, managers gain benefit from ensuring that investors choose the larger value of $k \in \{k_h^*, k_l^*\}$. So the managers decision could be determined by either maximising investors' expected utility of consumption or by looking at the metric;

$$M = \frac{k_l^* - k_h^*}{k_h^*} \quad (52)$$

In the latter case, if M is positive (negative) then the optimal choice for the manager is to choose e_l (e_h) = 0. It is *postulated* that these two methods for determining whether managers should set e_l or $e_h = 0$ will give the same result. This has been surprisingly awkward to prove as the intuition is clear. An investor will only choose to put more money into the investment process in one case than the other if it is more attractive in that case. The following postulate was checked for all simulations run below and was never violated but due to the general difficulty of providing analytical solutions in this environment, has not been formally proven.

Postulate 1 *If $M >, < 0$ then the expectation taken at time zero of investors' future utility of consumption is maximised by setting $e_l = 0, e_h = 0$ respectively. If $M = 0$ then investors are indifferent to whether e_h or $e_l = 0$*

On the grounds that managers choose the investment policy, this chapter uses M as the metric for deciding whether e_l or e_h should equal zero. This is also a much easier number to interpret as it is the relative difference in expected size of the firm with $e_l = 0$ rather than $e_h = 0$. Relative change in

expected utility has no meaning since utility can be arbitrarily rescaled and absolute change in expected utility is difficult to interpret. Finally, as with the example in the introduction, it will be shown that changes in expected utility are small in all cases but the value of M can be large — that is, managers have more incentive to get the decision right than shareholders.

The complexity in this model comes from having trading. It is not possible, in general, to determine analytical form for p, l , although it will be possible to generate numerical solutions. In the next subsection $l = 0$ will be imposed and r will be made non-stochastic. Some stylised facts can be determined analytically in this case. This section provides understanding for the more complex simulations that follow.

23.2 No trade

In the more general formulation given in section 23.3 there is a market for trade open at $t = 1$ in the low state so that the unemployed can trade investment obligations with the employed. This market will operate after the level of investment and employment status has been revealed but before the investment is made. This section, which deals with an economy with certain returns, r , fulfils two aims. First, conditions where this market is required — that is, that trading will occur if the market opens — is presented. Second, the optimal strategy for revealing investment uncertainty should no market for trade exist (that is, if $l = 0$ is imposed) at $t = 1$ will be developed. It will be shown that \mathcal{A}' will determine whether any financial market is required at $t = 1$. The optimal policy for revealing investment information in the

absence of a market will depend upon the sign of \mathcal{P} .

23.2.1 Is a market required?

Trading occurs if different investors value the claims differently. Remember that we can use the Euler equation to find the price p at which each investor would trade the claim should a market exist. Trading will occur if and only if these values differ. In this case, the Euler equation is given by $p + \tilde{i} = E[\tilde{d}U'(c_2)/U'(c_1)]$ as the true price of the claim includes the investment obligation inherent in buying the asset. The following result holds:

Result 1 *If \mathcal{A}' has the same sign for all levels of consumption and $\tilde{i} > 0$ then trading will occur at $t = 1$ if and only if $\mathcal{A}' \neq 0$. If $\mathcal{A}' < (>)0$ then trading will occur such that an unemployed investor is a seller (buyer).*

Proof

Trading will occur if one class of investor values the claim at $t = 1$ differently from the other class of investor. So, trading will occur if and only if $\partial p / \partial y \neq 0$. If the partial derivative is positive (negative) across the domain then the employed (unemployed) will value the asset more highly than the other type of investor and so will purchase the asset. Therefore, to establish the result, it is necessary and sufficient to demonstrate that $\partial p / \partial y$ has the opposite sign to \mathcal{A}' and that this sign is constant for all c . As the price of the asset is given by $p = E[\tilde{d}U'(c_2)/U'(c_1)] - \tilde{i}$ and as everything is known

with certainty when trading occurs, $\partial p / \partial y = \tilde{d} \partial [U'(c_2) / U'(c_1)] / \partial y$. Simple algebraic manipulation shows that:

$$\frac{\partial}{\partial y} \left[\tilde{d} \frac{U'(c_2)}{U'(c_1)} \right] = \tilde{d} [\mathcal{A}(c_1) - \mathcal{A}(c_2)] \left[\frac{U'(c_2)}{U'(c_1)} \right]$$

Now the ratio of marginal utilities is strictly positive. As $(1 + r) > 0$, \tilde{d} has the same sign as \tilde{i} . If $\tilde{i} > 0$ then $c_2 > c_1$ so $\mathcal{A}(c_1) - \mathcal{A}(c_2)$ will be positive, zero or negative if $\mathcal{A}' <, =, > 0$ respectively. The result has thus been established.

QED

This result is not intuitively obvious. Remember that trading occurs once all uncertainty has been resolved. This dependency on changes in risk aversion is perhaps surprising. Given that, for power and quadratic utility, $\mathcal{A}' <, > 0$ respectively, it is clear that there will be trading at $t = 1$ with the unemployed selling and buying respectively. For exponential utility, with $\mathcal{A}' = 0$, any market at $t = 1$ is redundant.

23.2.2 Optimal managerial policy

Having shown the conditions under which trading will occur, the optimal investment signalling policy in an economy where there is no market at $t = 1$ is now determined. The manager is deciding whether to fully reveal investment policy in the low or the high state. The manager will choose this policy depending on which strategy maximises k . Invoking postulate 1, this

is equivalent to maximising $E_0[u(c_1, c_2)]$. Having just demonstrated the role that \mathcal{A}' plays, this will highlight the role that \mathcal{P} plays in proceedings.

Introduce some new notation. Let $d := i(1 + r)$. Define $\bar{c}_{x1} := y_{x1} - ki$, $\bar{c}_{x2} := y_{x2} + kd$ for $x = h, e, u$. That is, these \bar{c} s are the consumption by each category of investor if there is no market for trade and if there is no uncertainty in investment. The expected (at $t = 0$) utility of consumption will be expressed in terms of these \bar{c} s:

$$\begin{aligned} 2E_0[U(c_1, c_2)] &= E_0[U(\bar{c}_{h1} - ke_h) + U(\bar{c}_{h2} + (1 + r)ke_h)] \\ &\quad + (1 - s)E_0[U(\bar{c}_{e1} - ke_l) + U(\bar{c}_{e2} + (1 + r)ke_l)] \\ &\quad + sE_0[U(\bar{c}_{u1} - ke_l) + U(\bar{c}_{u2} + (1 + r)ke_l)] \\ &\approx U(\bar{c}_{h1}) + U(\bar{c}_{h2}) + (1 - s)[U(\bar{c}_{e1}) + U(\bar{c}_{e2})] + \\ &\quad s[U(\bar{c}_{u1}) + U(\bar{c}_{u2})] + 0.5k^2\sigma_{e_h}^2[U''(\bar{c}_{h1}) \\ &\quad + (1 + r)^2U''(\bar{c}_{h2})] + 0.5k^2\sigma_{e_l}^2[(1 - s)U''(\bar{c}_{e1}) \\ &\quad + sU''(\bar{c}_{u1}) + (1 - s)(1 + r)^2U''(\bar{c}_{e2}) + s(1 + r)^2U''(\bar{c}_{u2})] \end{aligned}$$

Now the terms in $U(c_{xt})$ do not depend on the state containing investment uncertainty. Therefore the impact of concentrating investment uncertainty in the low state rather than in the high state is captured in the U'' terms. It is clear from these terms that the key question is whether or not the investor is prudent (in the technical sense that was introduced in chapters 3 & 4 — $U''' > 0$)¹⁰⁰. If the investor is prudent (which will be true for both power and exponential utility) then the second derivative terms associated with the low state will be more negative than the second derivative terms associated with the high state. Therefore it is clear that if there are no markets and the investor is prudent then, for any given k , the investor is

¹⁰⁰If $\mathcal{A}' < 0$ then $\mathcal{P} > 0$ but that the implication does not run the other way as $\mathcal{A}' < 0 \equiv U''' > (U'')^2/U'$ which is stronger than $U''' > 0$, the condition for prudence.

happier to have investment uncertainty in the high state than the low state. Invoking postulate 1 it can be seen that the optimal policy for managers is to set $e_l = 0$ if $\mathcal{P} > 0$, $e_h = 0$ if $\mathcal{P} < 0$ and, if $\mathcal{P} = 0$ then there is indifference (both for managers and investors).

Consider the simple example given in the introduction. Let $y_h = 1$ and $y_u = 0.4$. Let $r = 0.05, s = 0.1$. Let $\tilde{i} \in \{0.1, 0\}$ with equal probability in one state and let $\tilde{i} = 0.05$ with certainty in the other state. In the introduction, it was mentioned that for power utility with $\gamma = 3$ that $k_h^* = 0.0518 < k_l^* = 0.0807$. This is what would be expected given that power utility is prudent. Consider also exponential utility — $U(c) = -e^{-ac}$ — which is prudent and quadratic utility — $U(c) = Ac - Bc^2$ for $c \in [0, A/2B]$ — which is not prudent and for which $\mathcal{A}' > 0$. Choose parameters a, A, B so that $\mathcal{A} = 3$ at $c = 1$ ($a=3, A=2.667, B=1$). For exponential utility $k_h^* = 0.099$ and $k_l^* = 0.113$ while for quadratic utility, $k_h^* = k_l^* = 0.11521$. This demonstrates the result proven above (and, indeed, is the first evidence supporting postulate 1).

To summarise, if investors have power or quadratic utility then there would be trading in a market at $t = 1$ if the state is low. For power (quadratic) utility the unemployed will be the sellers (buyers) in this market if \tilde{i} is positive. Any such market is redundant if utility is exponential. In the absence of a market at $t = 1$, quadratic utility investors are indifferent between which state contains the investment uncertainty. With power or exponential utility, investors would prefer the investment uncertainty to be concentrated in the high state and hence will choose to hold more of the

assets if this is the case.

23.3 With a financial market

23.3.1 Certain returns

Having provided some analytical results with $l = 0$, relax this assumption and allow trading in the market. Power, exponential and quadratic utility functions are all examined in this subsection. For the moment, returns to scale r remain certain, but this assumption is relaxed below. The prices and quantities of trades are calculated numerically. The results of this section were calculated using Pascal programmes with algorithms based heavily on the recipes given in Press, Flannery, Teukolsky and Vetterling (1989). The functions and procedures that have been taken from this source are identifiable by the names and section numbers given in the footnotes. One of the Pascal programmes, “Powermin.pas”, is given in appendix 26. This provides the results for power utility. Similar programmes were used for quadratic and exponential utility. The structure of the programme is now briefly described.

- Iterations were run for 40 economies with different parameter values chosen to describe the economy. For all three utility functions, parameters are chosen so that the coefficient of absolute risk aversion at $c = 1$ is generated from a rectangular distribution¹⁰¹ in $(1, 10)$. r and i are generated from rectangular distributions with range $(0.01, 0.1)$ and $(0.1, 0.2)$ respectively. $y_h = 1$, $y_u = 0.4$ and $s = 0.05$ throughout.

¹⁰¹Using function “ran3”, §7.1.

- Within each of the 40 economies, the numerical algorithm was run twice: once with investment uncertainty in the low state and once with investment uncertainty in the high state.
- For each of the 40 economies, 100 random numbers, ε_j with $j \in \{1..100\}$ were drawn from a normal distribution¹⁰² with mean 0 and standard deviation 0.05 (these 100 values are the same for all economies). Outlines were excluded by replacing any ε_j that had higher absolute value than the lowest value of i for the 40 economies. This means that investment is never negative. In order to ensure that the mean value of the investment deviation was equal to zero for each economy 200 values were used for investment shocks; the 100 values of ε_j and the 100 values of $-\varepsilon_j$. So, \tilde{i} can take any one of 201 values which will be denoted by $\tilde{i}_w, w \in \{0, .., 200\}$. Define $\tilde{i}_0 = i$, $\tilde{i}_w = i + \varepsilon_w$ for $w \in \{1, ..100\}$ and $\tilde{i}_w = i - \varepsilon_{w-100}$, for $w \in \{101, .., 200\}$.
- Remembering that trading occurs once \tilde{i} has been revealed, for each of the 201 values that \tilde{i} will take it is possible to calculate, using a numerical non-linear simultaneous equations solver¹⁰³, unique values of p, l that satisfy the two Euler equations for any given k in each economy. That is, the price and quantities traded are determined by:

$$p - \tilde{i} = \frac{U'(c_{e2})}{U'(c_{e1})} = \frac{U'(c_{u2})}{U'(c_{u1})}$$

¹⁰²Using function “gasdev”, §7.2.

¹⁰³Procedure “mnewt”, §9.6

where consumption is defined by equations 49, 50. Notice that this is an approximation as the Euler equation gives the equilibrium price for an infinitesimal volume trade in equilibrium. In this case, l will be finite and so it is not clear that the true price of trade is exactly equal to p . This assumption, though, is assumed to be immaterial.

- k_h^*, k_l^* can now be calculated by using equation 51. c_{xt} for $x \in \{h, e, l\}, t \in \{1, 2\}$ are defined in equations 49, 50 with $\tilde{i} = \tilde{i}_0$ if investment is certain in state x and \tilde{i} takes one of the values of $\tilde{i}_w, w \in \{1, \dots, 200\}$ if investment is uncertain in state x . In this case:

$$E_{\tilde{i},r}[U(c_{xt})] = \begin{cases} U(c_{xt}) & e_x = 0 \\ \frac{1}{200} \sum_{\tilde{i}=\tilde{i}_1, \dots, \tilde{i}_{200}} U(c_{xt}) & e_x \neq 0 \end{cases} \quad (53)$$

- Finally, M , as given in equation 52, is calculated. If M is positive (negative) then managers will prefer to have investment certainty in the low (high) state.

The algorithms were checked for exponential and quadratic utility by re-running the programmes and imposing $p = l = 0$ in all cases rather than using the non-linear simultaneous equation solver to determine them endogenously. We know that investors with exponential utility find a financial market redundant and therefore M should be the same for each of the 40 economies whether we impose $p = l = 0$ or not in this case. For quadratic utility and no trading, investors should be indifferent between which state contains investment uncertainty and so M should be zero in each case. The

algorithms passed both of these tests. The non-linear simultaneous equations solver was checked separately to ensure that it was providing accurate values for p and l .

The results from running these simulations are what we might expect given the discussions above. For each of the 40 economies, M is always positive with power and exponential utility and always negative with quadratic utility. The average value of M was 5.5%, 3.4% and -2.5% for power, exponential and quadratic utility respectively. These are significant differences. Postulate 1 was not violated for any simulation. That is, in all 40 cases, the expected utility of consumption with $e_l = 0$ was higher for power and exponential utility and lower for quadratic utility than with $e_h = 0$. However, the utility functions were very flat close to the optimal and the change in expected utility for the investors from choosing $e_l = 0$ rather than $e_h = 0$ was very small. So, it is the manager who derives the greatest benefit from making the optimal investment choice.

In the two graphs in figure 11 it is shown how M varies with s (probability of unemployment), y_u (unemployment benefit), r (certain returns to scale), γ (coefficient of relative risk aversion) for power utility. These graphs were generated using a very similar algorithm to that described above. The main differences are that, in this case, 200 economies were used but only 5 values of ε_j were drawn in each case (so \tilde{i} takes one of 11 values). Also, rather than picking parameters at random, s, y_u, r, γ are examined across a grid. From these graphs it is clear that the two key variables in determining the

magnitude of M are the level of unemployment income and the coefficient of relative risk aversion. As expected, the lower the subsistence income in the low state and the higher the risk aversion the higher is M . The higher the probability of unemployed, the greater M , although this effect is not as strong as the effect of y_u . That is, if unemployment benefit is sufficiently low then the probability of becoming unemployed is largely irrelevant — it is just the possibility of potential unemployment that drives M . There is no clear relationship between the return to capital and M . This absence of any relationship between these two variables is confirmed by regressing M on r : the coefficient is very close to zero and is statistically highly insignificant.

23.3.2 State independent risky returns

An alternate version of the algorithm was run in which returns were risky but state independent. In the model described above returns are certain and trading occurs in the financial market when all information is known. The adjustment was to let the return to capital be stochastic (\tilde{r}) with \tilde{r} taking one of two values with equal probability $\tilde{r} = r \pm 0.1$. The true value of \tilde{r} is not revealed until $t = 2$: that is, after trading has occurred. In this case 40 economies were simulated with 20 values drawn for ε_j , so $\tilde{i}_w = \varepsilon_j$, $w \in \{1, \dots, 20\}$ and $\tilde{i}_w = -\varepsilon_{w-20}$, $w \in \{21, \dots, 40\}$. Define $r_1 := r - 0.1$, $r_2 := r + 0.1$. k_h^*, k_l^* can now be calculated by using equation 51. c_{xt} for $x \in \{h, e, l\}$, $t \in \{1, 2\}$ are defined in equations 49, 50 with $\tilde{i} = i$ if investment is certain in state x and \tilde{i} takes one of the values of \tilde{i}_w , $w \in \{1, \dots, 40\}$ if investment is

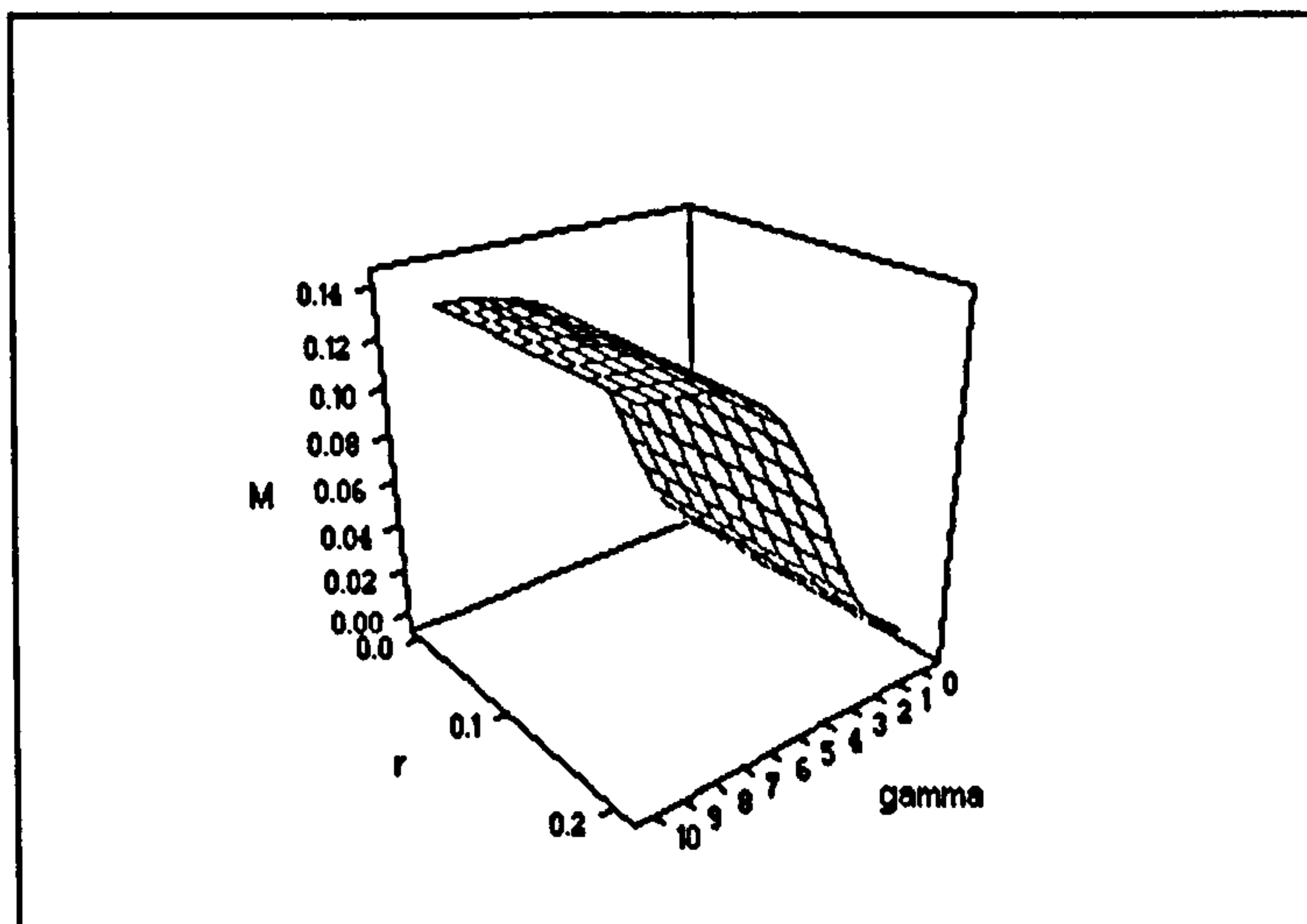
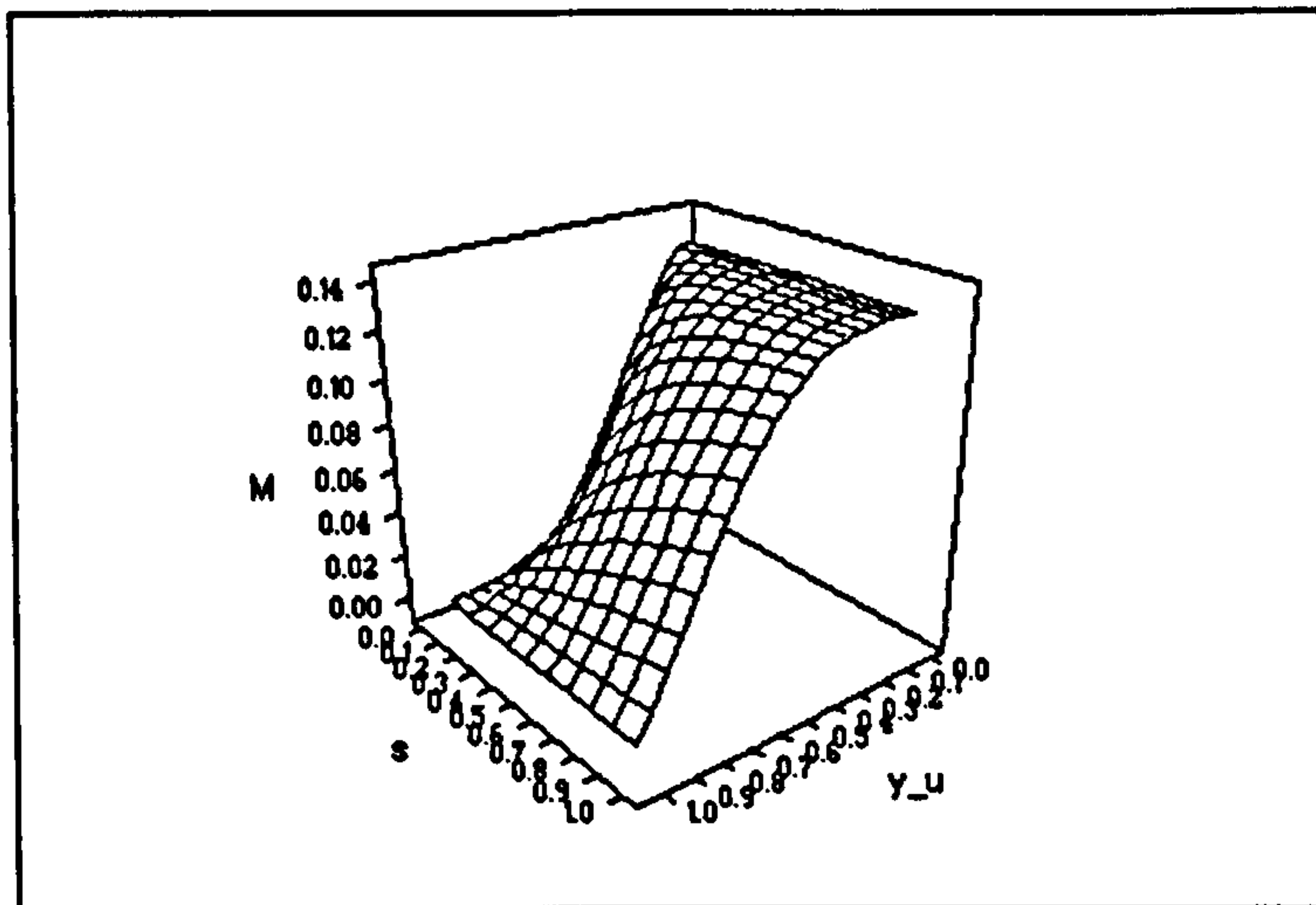


Figure 11: A graph showing how the significance of state dependent investment uncertainty varies with unemployment risk, unemployment income, risk aversion and fixed return on capital for power utility. M measures the additional percentage of shares investors will choose to take up at $t = 0$ if investment is uncertain in the high state rather than the low state. γ is the coefficient of relative risk aversion, r is the return on capital (non-stochastic). s is the unemployment rate in the low state and y_u is endowment in the unemployed state. For all iterations, $i = 0.1, y_h = 1$. For the top graph $\gamma = 3, r = 0.1$. For the bottom graph, $s = 0.05, y_u = 0.4$.

uncertain in state x . $\tilde{d} = (1 + r_1)\tilde{i}$ or $\tilde{d} = (1 + r_2)\tilde{i}$ with equal probability in either case. In this case:

$$E_{\tilde{i},r}[U(c_{xt})] = \begin{cases} \frac{1}{2} \sum_{r=r_1, r_2} U(c_{xt}) & e_x = 0 \\ \frac{1}{80} \sum_{r=r_1, r_2} \sum_{\tilde{i}=\tilde{i}_1, \dots, \tilde{i}_{40}} U(c_{xt}) & e_x \neq 0 \end{cases} \quad (54)$$

The following characteristics were observed. First, the sign of M for all 40 economies with each of the three utility functions was the same with risky returns as with certain returns. However, the absolute value of M was reduced. For power utility, the average value of M^{104} changed from 4.66% to 3.83% when returns were allowed to be risky. The analogous figures for exponential and quadratic utility were 2.89% to 2.39% and -2.14% to -1.76% respectively. For all 40 economies and for all three utility functions, the absolute value of M was lower when returns were risky rather than certain. Obviously, in each case as investors are risk averse, the utility of expected consumption and k were lower with risky returns than certain returns. Postulate 1 held in all cases. Simulations also confirmed that with state dependent risky returns, the financial market remains obsolete for exponential utility and if we force no trade ($l = 0$) for quadratic utility $M = 0$ in all cases.

23.3.3 State dependent risky returns

Finally, an algorithm was run for power utility where returns were risky and state dependent. That is, $\tilde{r} = r + \vartheta$ in the high endowment state and

¹⁰⁴This value of M is not directly comparable with that given above as the number of simulations run here is less than in the previous case.

$\tilde{r} = r - \vartheta$ in the state where unemployment is a possibility with $\vartheta > 0$ as unemployment is procyclical. In this case, as the state does not change from $t = 1$ to $t = 2$, returns are known when trading occurs but not at $t = 0$, when k is chosen. Another adjustment was made to the models described above. So far, it has been assumed that there is no real investment before $t = 1$ and no real investment after $t = 2$. However, it is reasonable to suppose that a company would be more able to finance new projects from internally generated funds at times of low unemployment. Therefore, the expression for c_{h1} in equation 49 is amended to $c_{h1} = y_h - k(\tilde{i} - \nu)$ where $\nu \geq 0$ represents the funding for the new project that comes from internally generated profits. In the low state, it is assumed that no such funding is available. So equations 49, 50 are amended in this instance to:

$$\begin{aligned}
c_{h1} &= y_h - k(\tilde{i} - \nu) \\
c_{e1} &= y_h - Slp - (k + Sl)\tilde{i} \\
c_{u1} &= y_u + pl - (k - l)\tilde{i} \\
c_{h2} &= y_h + k\tilde{i}(1 + r + \vartheta) \\
c_{e2} &= y_h + (k + Sl)\tilde{i}(1 + r - \vartheta) \\
c_{u2} &= y_u + (k - l)\tilde{i}(1 + r - \vartheta)
\end{aligned} \tag{55}$$

The algorithm used to calculate M in this case was identical to the one used to generate the two graphs in figure 11. Again, 200 points are chosen across a grid in this space and 5 values are chosen for ε_j . In this case $i = 0.15$ (not 0.1 as with the other two graphs), $y_u = 0.4$, $y_h = 1$, $\gamma = 3$, $r = 0.1$, $s = 0.05$. In figure 12 it is shown how M varies with ϑ, ν . It can be seen that M is monotonic increasing in ν , but, as ϑ gets greater so M gets smaller. This is consistent with the state independent risky returns result. Postulate

1 held for all 200 points.

The intuition is straightforward. As ν increases so the asset becomes more attractive (the company is expected to fund more of the investment itself) and k_l^*, k_h^* increase.

However, as ϑ increases the asset becomes more risky so less attractive causing k_l^*, k_h^* to decrease. In the first case investors have been persuaded to take on more assets because of a potential benefit should the state be high at $t = 1, 2$. However, if the state turns out to be low, then the effects of suboptimal investment are amplified. This drives M up. In the second case, the profit risk partially “dominates” the investment risk.

Concluding, in the case of state dependent returns, there are two counteracting effects. First, the variability in profits makes investors wary of the asset to start with. The effect of the additional risk caused by investment uncertainty is lowered. The offsetting effect is that companies are less able to finance projects from retained profits in the low state. There is a potential double hit in this case — investment may be “too high” and profits may be “too low” to fund it — causing a greater shock to investors’ cashflows. This issue is examined in a multiperiod context with no trading below.

24 A multi-period economy

In the previous section, a two-period model with trading was examined. In this section, it is assumed that trading is prohibited in order to convert to a multiperiod world. Subsection 24.1 considers a model very similar to that

given in the previous section. By transferring to a multiperiod economy the role that retained profits plays in the investment process becomes clearer (that is, ν of the previous section becomes endogenised). Subsection 24.2 then shows the advantage of a dividend smoothing model within an economy with endowment shocks. In the multiperiod case, only power utility with $\beta = 1$ is examined. There is one other important difference between the two period world and the economy presented here. In the former case, all investment was made at $t = 1$ and there was no investment after $t = 2$. In the multiperiod case it is assumed that there exists capital in the firm at $t = 0$ and there will continue to be capital in the firm after $t = T$. It is fluctuations in capital over $[0, T]$ that drive the model. This is more in keeping with an economy with long lived firms and shorter lived investors.

24.1 Certain and state dependent risky returns

This subsection aims to be the multiperiod equivalent of the two period cases of certain returns and state dependent risky returns given above. The economy develops as follows:

- At $t = 0$ the economy is in the high state. There is one company that announces that, for every claim that an investor wishes to purchase, an external source will place capital K_0 in a real investment process on the understanding that, at time T this source will then withdraw $K_T = K_0$ per claim. At each point t in between it will have capital employed $K_0 + e_{xt}$ per claim where e_{xt} is a mean zero random variable

whose variance is state dependent. If retained capital is not sufficient to satisfy $K_0 + e_{xt}$ then investors will be expected to meet the shortfall but, in return, investors will receive a dividend of any excess capital. As before, there are two states: an upstate and a downstate. The managers can choose $e_{lt} = 0 \forall t$ or $e_{ht} = 0 \forall t$ but not both. They announce this decision to investors at $t = 0$.

- In the interval $t \in \{1, \dots, T\}$ the economy is described by a Markov switching model.
- At time t , the economy will be in the same state as $t-1$ with probability $(1-q)$ and switch states with probability q . The return on capital over the period $[t-1, t]$ is given by $r_x \in \{r_h, r_l\}$ where x refers to the state at time t . So, if the economy is in the high state at t then it makes profits $r_h K_{t-1}$, per claim else profits are $r_l K_{t-1}$ per claim with $r_h \geq r_l$. r_h, r_l are known at $t = 0$. e_{xt} is then drawn and the company calculates the capital employed $K_t = K_0 + e_{xt}$ that it needs per claim for the next time period. So, the net payout to investors per claim is $d_t = r_x K_{t-1} - e_{xt}$, which can be negative.
- Investors receive endowment at t of $y_t = y_h$ if the state is high. If the state is low then a proportion s receive endowment $y_t = y_u$, the remainder $y_t = y_h$ with $y_h \gg y_u$. In the low state at any time point, the probability of becoming unemployed is not dependent on the previous employment status of the individual — it is s for each individual for

each period in which the economy is in the low state. That is, there is no persistence in unemployment in this model.

- Trading is prohibited.
- At time T , in order to ensure that the level of capital employed by the company is indeed K_T , it is necessary to impose the restriction that $e_{xT} = 0$ whatever state prevails at that time. The economy then closes at time T .
- At time $t = 0$ investors are given the choice of how many claims k they wish to own. Investors are *ex-ante* homogeneous and so choose the same k . If $e_{lt} = 0 \forall t$ then denote by k_l^* the number of claims chosen. Similarly, if $e_{ht} = 0 \forall t$ then k_h^* is the optimal number of shares. These are defined by

$$\begin{aligned} k_h^* &:= \max_{k|e_{ht}=0 \forall t} \sum_{t=1}^T E_0[U(y_t + kd_t)] \\ k_l^* &:= \max_{k|e_{lt}=0 \forall t} \sum_{t=1}^T E_0[U(y_t + kd_t)] \end{aligned} \quad (56)$$

- As in the two period case, the metric $M := (k_l^* - k_h^*)/k_h^*$ is used to determine the optimal managerial policy. Postulate 1 was not violated for any of the economies considered.

Simulations were run for 24 economies for both models. For each of these economies 250 simulations were run (denoted by σ). A numerical optimiser was used¹⁰⁵ to determine k_l^*, k_h^* as defined above. The state at time t varied

¹⁰⁵Again “brent”.

between simulations and so values of d_t also varied between simulations. Use d_t^σ to denote the dividend paid at time t in simulation σ . Expected utility of consumption was calculated as follows:

$$E_0[U(y_t + kd_t)] := \frac{1}{250} \sum_{\sigma=1}^{250} U^\sigma(c_{xt}) \quad (57)$$

$$U^\sigma(c_{xt}) := \begin{cases} U(y_h + kd_t^\sigma) & x = h \\ sU(y_l + kd_t^\sigma) + (1-s)U(y_h + kd_t^\sigma) & x = l \end{cases}$$

Define the variables. $\gamma = 3, y_h = 1, T = 100, q = 0.4$. The distribution of e_{xt} is either zero with certainty or rectangular in $[-0.25K_0, 0.25K_0]$. $K_T = K_0 = 0.2, s = 0.1$. $r_l \in \{-0.01, r_h\}$. $r_h \in \{0.01, 0.02, 0.03\}$ if $r_l = r_h$, otherwise $r_h \in \{0.03, 0.05, 0.07\}$. $y_u \in \{0.25, 0.5, 0.75, 1\}$. The results are presented in table 6. In the cases when k_h^* are positive¹⁰⁶ the values of $M, \Delta U$ are positive in all cases. *Investors and managers prefer investment uncertainty in states with low risk to personal capital.* Further, M is greater for risky r than certain r with the same expectation. This means that the possibility of low profits and a positive investment shock combined in the low state increases the aversion of individuals to the equity. Postulate 1 holds in all cases. The low value of ΔU should be emphasised. While utility can be arbitrarily rescaled, it is worth noting that the value for expected utility was in the $(-90, -40)$ range for all simulations. Therefore, the manager appears to have more to gain than the investor from making the optimal choice. This is one of the main reasons for using M as the main metric in this chapter.

¹⁰⁶Selling short has little meaning in this context as it is real investment opportunities that are at issue.

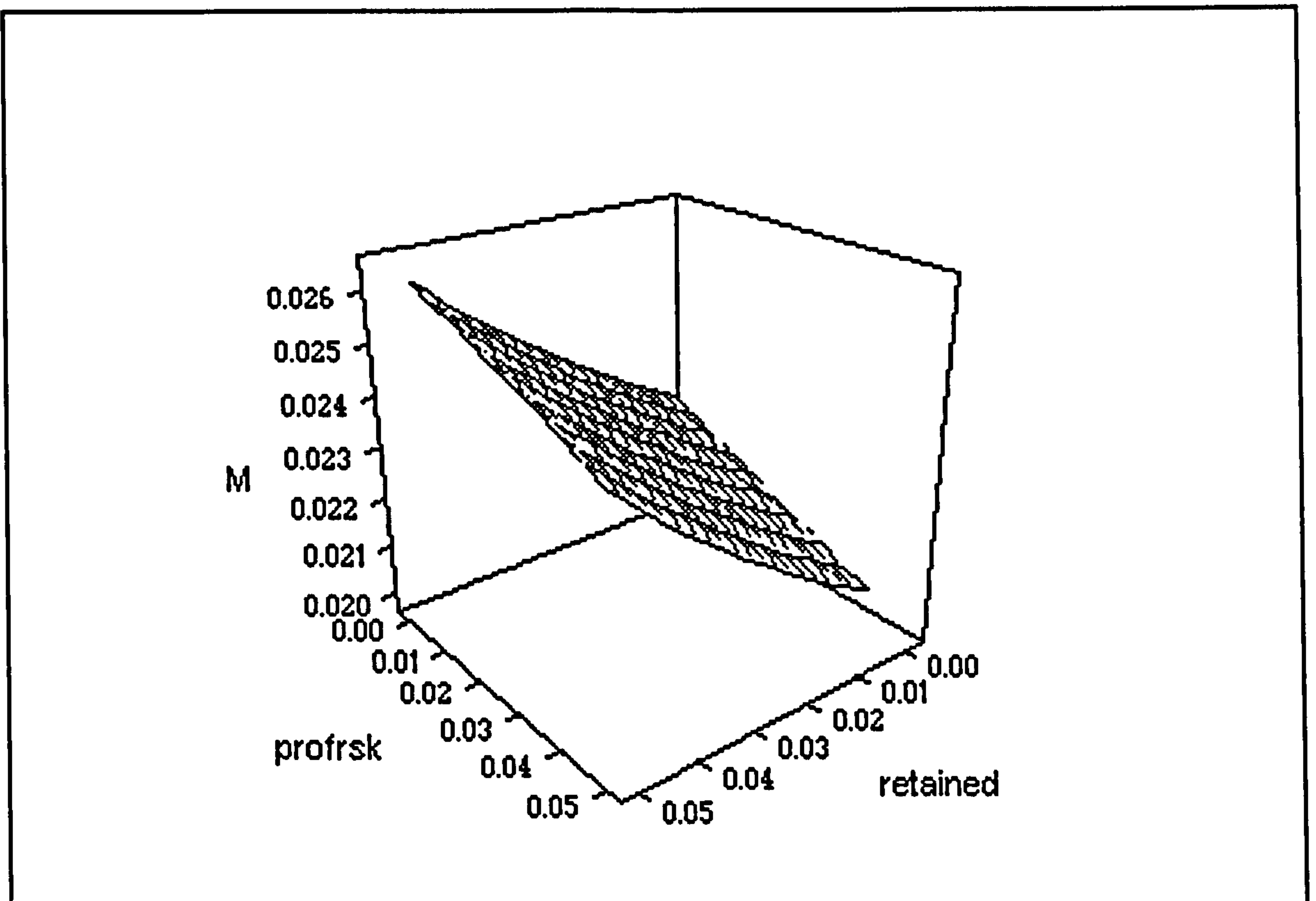


Figure 12: A graph showing how the significance of state dependent investment uncertainty varies with “profnsk” $\equiv \vartheta$, which is half the difference between returns in the low state and returns in the high state and “retained” $\equiv \nu$, which is the amount of investment funded from retained profits. M measures the additional percentage of shares that investors will choose to take up at $t = 0$ if investment uncertainty is in the high state rather than the low state. For all iterations, $i = 0.15, \gamma = 3, y_u = 0.4, s = 0.05, y_h = 1, E[r] = 0.1$.

| y_u | $r_l = r_h$ | | | | | $r_l = -0.01$ | | | | |
|-------|-------------|---------|---------|------|------------|---------------|---------|---------|------------|------------|
| | r_h | k_h^* | k_l^* | M | ΔU | r_h | k_h^* | k_l^* | M | ΔU |
| 0.25 | 0.01 | 0.15 | 0.56 | 2.73 | 0.134 | 0.03 | -0.09 | -0.29 | <i>n.a</i> | <i>n.a</i> |
| 0.5 | 0.01 | 0.48 | 0.85 | 0.78 | 0.043 | 0.03 | 0.22 | 0.40 | 0.80 | 0.003 |
| 0.75 | 0.01 | 0.72 | 0.84 | 0.17 | 0.011 | 0.03 | 0.60 | 0.71 | 0.18 | 0.007 |
| 1.0 | 0.01 | 0.79 | 0.82 | 0.04 | 0.003 | 0.03 | 0.76 | 0.80 | 0.05 | 0.003 |
| 0.25 | 0.02 | 0.30 | 1.06 | 2.52 | 0.610 | 0.05 | -0.05 | -0.15 | <i>n.a</i> | <i>n.a</i> |
| 0.5 | 0.02 | 0.95 | 1.67 | 0.75 | 0.187 | 0.05 | 0.56 | 0.99 | 0.78 | 0.062 |
| 0.75 | 0.02 | 1.42 | 1.65 | 0.17 | 0.049 | 0.05 | 1.20 | 1.45 | 0.21 | 0.042 |
| 1.0 | 0.02 | 1.56 | 1.62 | 0.04 | 0.011 | 0.05 | 1.45 | 1.57 | 0.09 | 0.018 |
| 0.25 | 0.03 | 0.45 | 1.51 | 2.34 | 1.330 | 0.07 | -0.01 | -0.02 | <i>n.a</i> | <i>n.a</i> |
| 0.5 | 0.03 | 1.41 | 2.44 | 0.73 | 0.412 | 0.07 | 0.86 | 1.54 | 0.79 | 0.165 |
| 0.75 | 0.03 | 2.10 | 2.44 | 0.16 | 0.107 | 0.07 | 1.72 | 2.16 | 0.26 | 0.107 |
| 1.0 | 0.03 | 2.30 | 2.38 | 0.04 | 0.025 | 0.07 | 2.04 | 2.32 | 0.14 | 0.059 |

Table 6: Multi-period economy with certain and state dependent risky returns. The number of assets chosen by investors k_h^*, k_l^* if investors are fully informed of the investment policy in the high state and low state respectively. M is the metric $(k_l^* - k_h^*)/k_h^*$, quoted only for positive k_h^* . ΔU , also quoted only for positive k_h^* , is the change in expected utility for investors from concentrating investment uncertainty in the high state rather than the low state. For all simulations $s = 0.1$, $q = 0.4$, $\gamma = 3$, $y_h = 1$, $T = 100$. These results are based on 250 simulations in each case.

24.2 A dividend smoothing model

The simulations in this chapter are concluded by presenting some results from one that demonstrates the merit of a dividend smoothing policy with rights issues concentrated in high states. In this section rights issues are more likely to happen in one state than another. Therefore, the expected level of investment is state dependent in this model. This, model, then, captures both the expected level of investment and uncertainty in investment and so is the closest model that is presented to the “real world” situation although trading is still prohibited. The benefits of concentrating rights issues in high endowment states is clearly demonstrated in this subsection. This confirms the intuition of the introduction that, by allowing the expected level of investment to be state dependent, the results will be strengthened and not weakened.

- At $t = 0$ the economy is in the high state. There is one company that states that, for every claim that an investor wishes to purchase, an external source will place capital K_0 in a real investment process on the understanding that, at time T this source will then withdraw $K_T = K_0$ per claim. At each point t in between the company wishes to have capital employed $K_{xt}^* = K_0 + e_{xt}$ per claim where e_{xt} is a mean zero random variable whose variance is state dependent. In this case, it is predetermined that $e_{lt} = 0 \forall t$ and e_{ht} is drawn from a rectangular distribution in $[-0.25K_0, 0.25K_0]$. This is known by investors at $t = 0$. In this case, though, the capital employed at time t , K_t need not always

equal K_t^* . The only restriction is that the capital at T must equal K_T .

- In the interval $t \in \{1, \dots, T\}$ the economy is described by a Markov switching model.
- At time t , the economy will be in the same state as $t-1$ with probability $(1-q)$ and switch states with probability q . The return on capital over the period $[t-1, t]$ is given by $r_x \in \{r_h, r_l\}$ where x refers to the state at time t . $r_h \geq r_l$ and r_h, r_l are known at $t=0$.
- Investors receive endowment at t of $y_t = y_h$ if the state is high. If the state is low then a proportion s receive endowment $y_t = y_u$, the remainder $y_t = y_h$ with $y_h \gg y_u$. In the low state at any time point, the probability of becoming unemployed is not dependent on the previous employment status of the individual — it is s for each individual for each period in which the economy is in the low state. That is, there is no persistence in unemployment in this model.
- The dividend process is described as follow. The net payout to investors at time t , $d_t := D_t + D_t^s - R_t$. $D_1 := \zeta K_0$ where ζ is the chosen dividend yield. At times $t \in \{2, \dots, T-1\}$, $D_t := \text{Max}(D_{t-1}, \zeta K_{t-1})$. This may be interpreted by saying that dividends never drop but will be raised if possible. There is, though, a risk this way that capital employed will fall so that K_T can never be reached. This is counteracted through right issues. Define numbers $\psi_x < 1$ for both states x so that, in state x , if, in the absence of a rights issue $K_t < \psi_x K_t^*$ there will be a

rights issue of exactly the required size to make $K_t = K_t^*$. Finally, if capital employed is getting very much greater than K_t^* then a “special” dividend D_t^s will be paid and investors will understand that this is a one-off payment. If, in the absence of this special dividend, $K_t > 1.1K_t^*$ then a special dividend will be paid to return K_t to K_t^* . Otherwise the special dividend is zero. $K_t = (1 + r_{xt})K_{t-1} - D_t - D_t^s + R_t$.

- Trading is prohibited.
- In order to ensure that capital truly equals K_T at time T , $D_T := D_{T-1}$, $D_T^s := 0$ and $R_T := K_T - (1 + r_{xT})K_{T-1} + D_T$.

The optimal value of k , denoted by k_s^* was calculated using equations 56, 57 above. 250 simulations were run for four economies. The pair $(\psi_h, \psi_l) \in \{(0.8, 0.7), (0.9, 0.8), (0.8, 0.9), (1, 1)\}$. Other variables are the same as for the ninth column, second row of table 6: $\gamma = 3, y_h = 1, T = 100, q = 0.4, s = 0.1, \zeta = 0.01\%$ (dividend yield), $K_0 = K_T = 0.2, r_h = 0.03, r_l = -0.01, y_u = 0.5$. The results can be directly compared with the value of k_l^* in the second row, ninth column of table 6 where $k_l^* = 0.396$.

When $(\psi_h, \psi_l) = (0.8, 0.7), (0.9, 0.8), (0.8, 0.9), (1, 1)$, $k_s^* = 1.166, 0.839, 0.366, 0.432$ respectively. Defining ΔU in this case to be the difference in expected utility between the model that generated $k_l^* = 0.396$ and the dividend smoothing policy, $\Delta U = 0.154, 0.084, -0.009, 0.004$ respectively. So, again, proposition 1 holds. Both managers and investors prefer this dividend smoothing policy to the investment smoothing policy given in the previous

section if $\psi_h > \psi_l$. In this case, dividends are always paid in the low state and so consumption is, on average, higher in these states. There is sometimes recourse to rights issues, but as these are concentrated in the high states this acts as a consumption smoothing effect. So, the number of claims chosen by investors may be around three times greater if we allow the expected value of investment to be state dependent than if we do not. Notice that rarer, but more severe, rights issues are preferred to smaller more regular rights issues: that is k^* is greater with $(\psi_h, \psi_l) = (0.8, 0.7)$ than $(0.9, 0.8)$. This is because, in the former case, small shortfalls in capital can often be reclaimed by future profits without necessarily having a rights issue.

25 Conclusion

In this chapter an economy has been examined where there is personal endowment risk and all financial investment is in the equity of a firm that sometimes invests suboptimally and does not fully reveal its investment plans. The economy has been constructed in such a way that there are no agency problems: a benevolent manager and a selfish manager always make the same choices. There are two effects at work. First, investors want *positive* endowment shocks in states where there is low risk to personal capital. Second, the simulations show that investors are also more willing to absorb *mean zero* investment shocks in this state as well. Therefore, it is contested that managers should amend any prior investment plans in such a way that dividends are not cut and rights issues are not undertaken in states where there is a high

risk to individual endowment. Dividends should be smoothed in order to inform investors of the net payoffs that will come from their portfolio in low states and rights issues should be concentrated in high states for companies to fulfil their investment plans. The two-period model of section 23 shows that even with the introduction of a financial market, this is still the optimal corporate finance policy. This is, in some ways, similar to the overinvestment model of Jensen (1986) except that here there is no systematic tendency to overinvest and agency issues are not needed to motivate the model.

While no existing empirical study can be considered a direct test of this model, the balance of evidence from this literature is broadly supportive of the model. Further, it has been argued that the assumptions needed to generate the results are reasonable and have, certainly to this author, more intuitive appeal than some existing theories of optimal dividend policy. Much further work is needed, both theoretical and empirical, but it is contested that this model merits serious evaluation.

26 Appendix

```
{Powermin.Pas}
CONST
np = 2;
runs=100;
runs2=40;
TYPE
RealArrayNP = ARRAY [1..np] OF real;
RealArrayNPbyNP = ARRAY [1..np,1..np] OF real;
IntegerArrayNP = ARRAY [1..np] OF integer;
VAR
{These are the variables for the preset functions/procedures}
mcf1,idum: integer;
mcf2: RealArrayNP;
mcf3,GasdevIset,Ran3Inext,Ran3Inextp: integer;
mcf4,mcf5,GasdevGset: real;
Ran3Ma: Array[1..55] of real;
```

```
    {These are the variables for my bit}
r,mu,gamma,bx1: array[1..runs2] of real;
err: array[1..runs] of real;
hh,jj,kk,mm,startup: integer;
BB,yu,ye,s,Es,d,dcert,i,claims,utilt2: real;
cert,min1,max1,avg1,return,ax1,cx1: real;
fileA,fileB,fileC,fileD: text;
checker: string;
```

```
    FUNCTION ran3(VAR idum: integer): real;
CONST
mbig = 4.0e6;
mseed = 1618033.0;
ms = 0.0;
fac = 2.5e-7;
VAR
i,ii,k: integer;
mj,mk: real;
BEGIN
IF idum<0 THEN BEGIN
mj:=mseed+idum;
IF mj >=0.0 THEN
mj:=mj-mbig*trunc(mj/mbig)
ELSE
mj:=mbig-abs(mj)+mbig*trunc(abs(mj)/mbig);
Ran3Ma[55]:=mj;
mk:=1;
FOR i:=1 to 54 DO BEGIN
ii:=21*i MOD 55;
Ran3Ma[ii]:=mk;
mk:=mj-mk;
IF mk<ms THEN mk:=mk+mbig;
mj:=Ran3Ma[ii]
END;
FOR k:=1 TO 4 DO BEGIN
FOR i:=1 TO 55 DO BEGIN
Ran3Ma[i]:=Ran3Ma[i]-Ran3Ma[1+((i+30) MOD 55)];
IF Ran3Ma[i]<ms THEN Ran3Ma[i]:=Ran3Ma[i]+mbig
END
END;
Ran3Inext:=0;
Ran3Inextp:=31;
idum:=1
END;
Ran3Inext :=Ran3Inext+1;
IF Ran3Inext=56 THEN
Ran3Inext:=1;
Ran3Inextp := Ran3Inextp + 1;
IF Ran3Inextp = 56 THEN Ran3Inextp:=1;
mj:=Ran3Ma[Ran3Inext] - Ran3Ma[Ran3Inextp];
IF mj<ms THEN mj:=mj+mbig;
Ran3Ma[Ran3Inext]:=mj;
ran3:=mj*fac
END;
```

```
    FUNCTION gasdev(VAR idum:integer): real;
VAR
fac,r,v1,v2: real;
BEGIN
IF GasdevIset = 0 THEN BEGIN
REPEAT
v1:=2.0*ran3(idum)-1.0;
v2:=2.0*ran3(idum)-1.0;
r:=sqr(v1)+sqr(v2);
UNTIL (r<1.0) AND (r>0.0);
fac:=sqrt(-2.0*ln(r)/r);
GasdevGset:=v1*fac;
```

```

gasdev:=v2*fac;
GasdevIset:=1
END
ELSE BEGIN
GasdevIset:=0;
gasdev:=GasdevGset;
END
END;

PROCEDURE usrfun(VAR x: RealArrayNP;
n: integer;
VAR alpha: RealArrayNPbyNP;
VAR beta: RealArrayNP;
xs: real);
VAR
plocal,llocal,cu0local,cu1local,muulocal: real;
ce0local,ce1local,muelocal,XUlocal,XElocal:real;
VUlocal,VElocal : real;
BEGIN
plocal:=x[1];
llocal:=x[2];
IF hh=1 THEN BEGIN
cu0local:=yu+(llocal*plocal)-((xs-llocal)*i);
cu1local:=yu+(d*(xs-llocal));
ce0local:=ye-(Es*llocal*plocal)-((xs+(Es*llocal))*i);
ce1local:=ye+((xs+(Es*llocal))*d);
muulocal:=Exp(-gamma[jj]*ln(cu1local/cu0local));
muelocal:=Exp(-gamma[jj]*ln(ce1local/ce0local));
XUlocal:= 1-((d*llocal*gamma[jj]*muulocal)/cu0local) ;
XElocal:= 1+((d*Es*llocal*gamma[jj]*muelocal)/ce0local);
VUlocal:= -d*gamma[jj]*muulocal*((d/cu1local)+((plocal+i)/cu0local));
VElocal:= d*Es*gamma[jj]*muelocal*((d/ce1local)+((plocal+i)/ce0local));
beta[1] := -plocal-i+(d*muulocal);
beta[2] := -plocal-i+(d*muelocal);
END ELSE BEGIN
cu0local:=yu+(llocal*plocal)-((xs-llocal)*mu[jj]);
cu1local:=yu+(dcert*(xs-llocal));
ce0local:=ye-(Es*llocal*plocal)-((xs+(Es*llocal))*mu[jj]);
ce1local:=ye+((xs+(Es*llocal))*dcert);
muulocal:=Exp(-gamma[jj]*ln(cu1local/cu0local));
muelocal:=Exp(-gamma[jj]*ln(ce1local/ce0local));
XUlocal:= 1-((dcert*llocal*gamma[jj]*muulocal)/cu0local) ;
XElocal:= 1+((dcert*Es*llocal*gamma[jj]*muelocal)/ce0local);
VUlocal:= -dcert*gamma[jj]*muulocal*((dcert/cu1local)+((plocal+mu[jj])/cu0local));
VElocal:= dcert*Es*gamma[jj]*muelocal*((dcert/ce1local)+((plocal+mu[jj])/ce0local));
beta[1] := -plocal-mu[jj]+(dcert*muulocal);
beta[2] := -plocal-mu[jj]+(dcert*muelocal);
END;
alpha[1,1]:=XUlocal;
alpha[1,2]:=VUlocal;
alpha[2,1]:= XElocal;
alpha[2,2]:= VElocal;
END;

PROCEDURE ludcmp(VAR a: RealArrayNPbyNP;
n: integer;
VAR indx: IntegerArrayNP;
VAR d: real);

CONST
tiny = 1.0e-20;
VAR
k,j,imax,i : integer;
sum,dum,big: real;
vv: ARealArrayNP;
BEGIN
new(vv);
d:=1.0;
FOR i:=1 TO n DO BEGIN
big:=0.0;
FOR j:=1 TO n DO
IF abs(a[i,j]) > big THEN big:=abs(a[i,j]);
IF big=0.0 THEN BEGIN
writeln('pause in LUDCMP - singular matrix');
readln
END;
vv[i] :=1.0/big
END;
FOR j:=1 to n DO BEGIN
FOR i:=1 to j-1 DO BEGIN
sum:=a[i,j];
FOR k:=1 to i-1 DO
sum:=sum-a[i,k]*a[k,j];
a[i,j]:=sum
END;
big:=0.0;
FOR i:=j to n DO BEGIN
sum:=a[i,j];

```

```

FOR k:=1 to j-1 DO
sum:=sum-a[i,k]*a[k,j];
a[i,j]:=sum;
dum:=vv^i*abs(sum);
IF dum >=big THEN BEGIN
big:=dum;
imax:=i
END
END;
IF j<>imax THEN BEGIN
For k:=1 TO n DO BEGIN
dum:=a[imax,k];
a[imax,k]:=a[j,k];
a[j,k]:=dum
END;
d:=-d;
vv^i[imax]:=vv^i[j]
END;
indx[j]:=imax;
IF a[j,j]=0.0 THEN a[j,j]:=tiny;
IF j<>n THEN BEGIN
dum:=1.0/a[j,j];
FOR i:= j+1 to n DO
a[i,j]:= a[i,j]*dum
END
END;
dispose(vv)
End;

```

```

PROCEDURE lubksb(VAR a: RealArrayNPbyNP;
n: integer;
VAR indx: IntegerArrayNP;
VAR b: RealArrayNP);

```

```

VAR
j,ip,ii,i: integer;
sum: real;
BEGIN
ii:=0;
FOR i:=1 to n DO BEGIN
ip:=indx[i];
sum:=b[ip];
b[ip]:=b[i];
IF ii<>0 THEN
FOR j:=ii to i-1 DO
sum:=sum-a[i,j]*b[j]
ELSE IF sum<>0.0 THEN
ii:=i;
b[i]:=sum
END;
FOR i:=n DOWNT0 1 DO BEGIN
sum:=b[i];
FOR j:=i+1 TO n DO
sum:=sum-a[i,j]*b[j];
b[i]:=sum/a[i,i]
END
END;

```

```

PROCEDURE mnewt(ntrial: integer;
VAR x: RealArrayNP;
n: integer;
tolx,tolf: real;
xy: real);
LABEL 99;
VAR
k,i: integer;
errx,errf,d: real;
beta: ARealArrayNP;
alpha: ARealArrayNPbyNP;
indx: AIntegerArrayNP;
BEGIN
new(beta);
new(alpha);
new(indx);
FOR k:=1 to ntrial DO BEGIN
usrfun(x,n,alpha^,beta^,xy);
errf:=0.0;
FOR i:=1 to n DO
errf:=errf+abs(beta^i[i]);
IF errf <= tolf THEN GOTO 99;
ludcmp(alpha^,n,indx^,d);
lubksb(alpha^,n,indx^,beta^);
errx:=0.0;
FOR i:=1 to n DO BEGIN
errx:=errx+abs(beta^i[i]);
x[i]:=x[i]+beta^i[i]
END;
IF errx <= tolx THEN GOTO 99

```

```

END;
99:
dispose(indx);
dispose(alpha);
dispose(beta)
END;

FUNCTION func(xx: real): real;
VAR
cu1global,ce0global,ce1global,utilu,utile,utilt : real;
pglobal,lglobal,cu0global: real;
yh,probbhigh,ch0global,ch1global,utilh : real;
rr,ll: integer;
BEGIN;
avg1:=0;
yh:=1;
probbhigh:=0.5;
for kk:=1 to runs do begin
{Investment noise - gaussian with mean 0.03}
for ll:=1 to 2 do begin
If ll=1 then rr:=1 else rr:=-1;
i:=mu[jj]+(err[kk]*rr);
return:=r[jj];
d:=(1+return)*i;
dcert:=(1+return)*mu[jj];

mnewt(mcf1,mcf2,mcf3,mcf4,mcf5,xx);
pglobal:=mcf2[1] ;
lglobal:=mcf2[2];

If hh=1 THEN BEGIN
cu0global:=yu+(lglobal*pglobal)-((xx-lglobal)*i);
cu1global:=yu+(d*(xx-lglobal));
ce0global:=ye-(Es*lglobal*pglobal)-((xx+(Es*lglobal))*i);
ce1global:=ye+((xx+(Es*lglobal))*d);
ch0global:=yh-(xx*mu[jj]);
ch1global:=yh+(xx*dcert);
END ELSE BEGIN
cu0global:=yu+(lglobal*pglobal)-((xx-lglobal)*mu[jj]);
cu1global:=yu+(dcert*(xx-lglobal));
ce0global:=ye-(Es*lglobal*pglobal)-((xx+(Es*lglobal))*mu[jj]);
ce1global:=ye+((xx+(Es*lglobal))*dcert);
ch0global:=yh-(xx*i);
ch1global:=yh+(xx*d);
END;

utilu:=(Exp((1-gamma[jj])*ln(cu0global))+Exp((1-gamma[jj])*ln( cu1global)))/(1-gamma[jj]);
utile:=(Exp((1-gamma[jj])*ln(ce0global))+Exp((1-gamma[jj])*ln( ce1global)))/(1-gamma[jj]);
utilh:=(Exp((1-gamma[jj])*ln(ch0global))+Exp((1-gamma[jj])*ln( ch1global)))/(1-gamma[jj]);
utilt:=(probbhigh*utilh)+((1-probbhigh)*((s*utilu)+((1-s)*utile) ));

avg1:=avg1+utilt;
end;
end;
func:=-avg1/(2*runs); {As we are using a minimizer}
END;

FUNCTION func2(xx: real): real;
VAR
cu1global,ce0global,ce1global,utilu,utile,utilt : real;
pglobal,lglobal,cu0global: real;
yh,probbhigh,ch0global,ch1global,utilh : real;
rr,ll: integer;

BEGIN;

yh:=1;
probbhigh:=0.5;
return:=r[jj];
dcert:=(1+return)*mu[jj];
cu0global:=yu-(xx*mu[jj]);
cu1global:=yu+(dcert*xx);
ce0global:=ye-(xx*mu[jj]);
ce1global:=ye+(xx*dcert);
ch0global:=yh-(xx*mu[jj]);
ch1global:=yh+(xx*dcert);

utilu:=(Exp((1-gamma[jj])*ln(cu0global))+Exp((1-gamma[jj])*ln( cu1global)))/(1-gamma[jj]);
utile:=(Exp((1-gamma[jj])*ln(ce0global))+Exp((1-gamma[jj])*ln( ce1global)))/(1-gamma[jj]);
utilh:=(Exp((1-gamma[jj])*ln(ch0global))+Exp((1-gamma[jj])*ln( ch1global)))/(1-gamma[jj]);
utilt:=(probbhigh*utilh)+((1-probbhigh)*((s*utilu)+((1-s)*utile) ));

func2:=-utilt; {As we are using a minimizer}
END;

FUNCTION brent(ax,bx,cx,tol : real;
VAR xmin : real): real;
LABEL 99;

```



```

CONST
itmax = 100;
cgold = 0.3819660;
zeps = 1.0e-10;
VAR
a,b,d,e,etemp : real;
fu,fv,fw,fx : real;
iter : integer;
p,q,r,tol1,tol2 : real;
u,v,w,x,xm : real;

FUNCTION sign(a,b: real) :real;
BEGIN
IF b >= 0.0 THEN sign:= abs(a) ELSE sign := -abs(a)
END;
BEGIN
IF ax < cx THEN a:=ax ELSE a:=cx;
IF ax > cx THEN b:=ax ELSE b:=cx;
v:=bx;
w:=v;
x:=v;
e:=0.0;
if startup = 1 then fx:=func2(x) else fx :=func(x);
fv:=fx;
fw:=fx;
FOR iter :=1 TO itmax DO BEGIN
xm:=0.5*(a+b);
tol1:= tol*abs(x)+zeps;
tol2:=2.0*tol1;
IF abs(x-xm) <=tol2-0.5*(b-a) THEN GOTO 99;
IF abs(e) > tol1 THEN BEGIN
r:=(x-w)*(fx-fv);
q:=(x-v)*(fx-fw);
p:=(x-v)*q-(x-w)*r;
q:=2.0*(q-r);
IF q>0.0 THEN p:=-p;
q:=abs(q);
etemp:=e;
e:=d;
IF (abs(p) >= abs(0.5*q*etemp)) OR (p<=q*(a-x))
OR (p >= q*(b-x)) THEN BEGIN
IF x>=xm THEN e:=a-x
ELSE e:=b-x;
d:=cgold*e
END
ELSE BEGIN
d:=p/q;
u:=x+d;
IF (u-a < tol2) OR (b-u < tol2) THEN d:=sign(tol1,xm-x)
END
END
ELSE BEGIN
IF x >=xm THEN e:=a-x
ELSE e:=b-x;
d:=cgold*e
END;
IF abs(d) >=tol1 THEN u:=x+d
ELSE u:=x+sign(tol1,d);
if startup=1 then fu:=func2(u) else fu:=func(u);
IF fu <=fx THEN BEGIN
IF u>=x THEN a:=x ELSE b:=x;
v:=w;
fv:=fw;
w:=x;
fw:=fx;
x:=u;
fx:=fu
END
ELSE BEGIN
IF u < x THEN a:=u ELSE b:=u;
IF (fu <=fw) OR (w=x) THEN BEGIN
v:=w;
fv:=fw;
w:=u;
fw:=fu;
END
ELSE IF (fu <=fv) OR (v=x) OR (v=w) THEN BEGIN
v:=u;
fv:=fu
END
END
END;
writeln('pause in routine BRENT - too many iterations');
99:
xmin:=x;
brent:=fx
END;

```

```

BEGIN
assign(fileA, 'c:/thesis/chapter8/data/powermin.txt');
rewrite(fileA);
assign(fileB, 'c:/thesis/chapter8/data/powermi2.txt');
rewrite(fileB);
assign(fileC, 'c:/thesis/chapter8/data/powerchk.txt');
rewrite(fileC);

Gasdev1set:=0;
mcf1:=5000; {Number of iterations}
mcf2[1]:=0; {This is the price}
mcf2[2]:=0; {Number of shares sold}
mcf3:=2; {Number of unknowns}
mcf4:=1e-11; {Error size - 1}
mcf5:=1e-11; {Error size - 2}

{***** SET GLOBAL VARIABLES *****)

{Employed income} ye:=1;yu:=0.4;
{Percentage unemployed} s:=0.05; Es:=s/(1-s);
BB:=1;

idum:=-1;
for jj:=1 to runs2 do begin
gamma[jj]:=1+(ran3(idum)*9);
end;

{Level of return - rectangular in [0.01,0.1]}
idum:=-2;
for jj:=1 to runs2 do r[jj]:=0.01+(ran3(idum)*0.09);

{Expected investment - Rectangular in [0.1,0.2]}
min1:=9999;
idum:=-3;
for jj:=1 to runs2 do begin
mu[jj]:=0.1+(ran3(idum)*0.1);
if mu[jj] < min1 then min1:=mu[jj]
end;

max1:=-9999;
idum:=-4;
for jj:=1 to runs do begin
err[jj]:=99999;
while abs(err[jj]) > min1 do err[jj]:=gasdev(idum)*0.05;
if abs(err[jj]) > max1 then max1:=abs(err[jj])
end;
writeln('min1:=',min1,' max1:=',max1);

for hh:=1 to 2 do begin
for jj:=1 to runs2 do begin
writeln(hh, ' ', jj);

If hh=1 then begin
writeln('Starting');
startup:=1;
claims:=0.1;

bx1[jj]:=brent(0,0.1,1,1e-10,claims);
bx1[jj]:=claims;
writeln('bx1=',bx1[jj]);
end;

ax1:=bx1[jj]-0.01;
cx1:=bx1[jj]+0.01;

While func(ax1) < func(bx1[jj]) do begin
ax1:=ax1-0.01;
bx1[jj]:=bx1[jj]-0.01;
cx1:=cx1-0.01;
end;
While func(cx1) < func(bx1[jj]) do begin
ax1:=ax1+0.01;
bx1[jj]:=bx1[jj]+0.01;
cx1:=cx1+0.01;
end;
writeln('ax1:=',ax1,' bx1=',bx1[jj],' cx1:=',cx1);
If (func(bx1[jj])<func(ax1)) and (func(bx1[jj])<func(cx1))
then checker:='OK' else checker := 'Trouble';
writeln(checker);

startup:=2;
claims:=bx1[jj];
utilt2:=brent(ax1,bx1[jj],cx1,1e-15,claims);
writeln('claims=',claims);
writeln(fileC,hh,jj,bx1[jj],claims,checker);
bx1[jj]:=claims;

```

```
If hh = 1 then
writeLn(fileA,claims,utilt2);
If hh = 2 then
writeLn(fileB,claims,utilt2);
end;
end;
close(fileA);
close(fileB);
close(fileC);
END.
```

Part VI

Unemployment shocks and asset returns

Unemployment shocks and asset returns¹⁰⁷

Abstract

In an attempt to explain the long term average equity premium and real riskfree rate, several authors have modelled the effect of unemployment risk on predicted average asset returns. Theoretically these models are appealing since quite “small” uninsurable nonmarketable capital can have large effects on asset prices. Curiously, existing tests have concentrated on examining the static properties of asset returns (expected returns and variance / covariance characteristics) rather than the dynamic properties in such contexts. If unemployment is a key state variable determining the average equity premium and real riskfree rate then an empirical consequence is that unemployment shocks should be a major factor influencing variations over time in asset returns. Using UK and US data, this is the first study that examines changes in asset returns as unemployment risk varies. It is found that the riskfree rate rises prior to “bad” unemployment news, which is difficult to reconcile with the precautionary savings motive.

¹⁰⁷Versions of this chapter were presented to the Financial Options Research Center, University of Warwick, January 1995, the British Accounting Association Doctoral Colloquium, April 1995 and the Department of Accounting and Finance, Lancaster University, June 1995. I am grateful to the participants for their useful comments. I would also like to thank Jeremy Smith and Sanjay Yadav for their econometrics advice.

27 Introduction

The role that uninsurable shocks might play in explaining the Mehra & Prescott puzzles was discussed in depth in chapter 3. It was emphasised that the reason why, in the Weil (1992a) and Mankiw (1986) type models, uninsurable risk dramatically alters predicted average asset returns is that all risk is persistent within these economies. However, the multiperiod work of Heaton & Lucas (see, for example Lucas (1994), Heaton and Lucas (1995)) has estimated that real idiosyncratic endowment risk is not sufficiently persistent to dramatically alter the average real riskfree rate and equity premium. However, the cross-sectional data on the length of shocks to personal income does not provide accurate estimates. Therefore, while Heaton & Lucas make a significant contribution by showing how theoretically important the persistence of idiosyncratic endowment is in determining the testable implications of incomplete market models, their data does not provide conclusive evidence. This chapter aims to contribute to the debate on whether unemployment shocks are likely to lie at the heart of a valid explanation of Mehra & Prescott's puzzles.

Existing tests of the application of incomplete market theory to the Mehra & Prescott puzzles has concentrated on the *static* properties of asset returns. The Mehra & Prescott puzzles, Hansen-Jagannathan bounds tests and examinations of the (co)variance structure of asset returns discussed in chapter 3 do not consider how asset prices respond to changes in uninsurable risk. This chapter is based on the premise that if uninsurable risk is the key variable

determining the average equity premium and real riskfree rate then *changes* in the magnitude of uninsurable risk will be a key state variable influencing asset returns at the time. That is, this chapter is the first examination that considers how the static models discussed in chapter 3 can be tested in a dynamic framework.

The work to follow is best demonstrated by an example. Consider a one-period model different to, but in the spirit of, Mankiw (1986)¹⁰⁸. At $t = 0$ everyone is employed with consumption c_0 . At time $t = 1$ a percentage of the population s will become unemployed and receive endowment y_u while the remainder receive y_h . There is a riskfree asset in zero net supply and each investor holds one share in the risky asset so that $c_1 = y + d$. Here s is the key state variable, so that d is a function of s . Let $c_0 = 1, y_h = 0.8, y_u = 0.4$. Use power utility with $\beta = 1, \gamma = 3$. Let the probability of unemployment in the next period be 60% for 5% unemployment when $d = 0.3$ and 40% for 6% unemployment when $d = 0.2$. Notice that this is an entirely new class of model. Throughout the examination of incomplete market work in chapters 3,4 & 5 it was assumed that the risk to personal income is known with certainty at $t = 0$. The idea that not only are there low probability, high impact shocks to personal capital but also that it is also not possible to know exactly what the likelihood of this shock is in advance has not featured to date in the literature. It is easily verified that $p_f = 0.9769, p_m = 0.2443$ in this example. First consider the realised returns to the market over the interval

¹⁰⁸Independent dividend and labour income risk in the style of Weil (1992a) is not considered in this chapter.

$[0, 1]$. If unemployment turns out to be 5% then the realised (simple) return is 22.8%, while if unemployment turns out to be 6% then the realised return is -18.1%. So, the *ex-post* return to the market is negatively correlated with the level of unemployment as would be expected in a model where dividends drop when unemployment rises.

The *ex-post* return to the riskfree asset is, of course, fixed as the asset pays 1 at $t = 1$ whatever the state. Suppose, though, an instant after $t = 0$, “news” emerged so that investors’ now believe that there is a 75% probability that the unemployment level will be 5% in the next time period and a 25% probability that the unemployment level will be 6%. In this case, $p_f = 0.9166, p_m = 0.2445$. So, the reduced precautionary savings motive pulls down the price of the riskfree asset. There are two effects at work on the risky asset. First, the reduced desire to precautionary save makes all assets less desirable. This pressure forces the price p_m down. However, the expected higher dividend to be paid by the market next period drives the price up. So the impact of this news unambiguously pulls p_f down but the effect on p_m is more complicated. So, from this analysis there are two testable hypotheses: (i) if unemployment is lower (higher) at $t = 1$ than was expected at $t = 0$ then the *ex-post* return to the equity premium portfolio over the interval $[0, 1]$ will be higher (lower) than expected at $t = 0$. It is difficult to predict how the market will respond to news concerning unemployment at $t = 1$ over the interval $[0, 1]$ as the dividend and precautionary savings effects conflict with each other and (ii) as the true level of unemployment at $t = 1$

reveals itself over the interval $[0, 1]$ so the precautionary savings motive will drive changes in the riskfree rate. If unemployment is lower (higher) at $t = 1$ than expected at $t = 0$ then the riskfree rate will have risen (fallen) during the period.

This chapter examines the movements in asset returns prior to good and bad unemployment announcements. The key test will be to see whether the riskfree rate falls (rises) prior to poor (good) unemployment news, as the precautionary savings motive would imply. The *ex-post* excess return to the market index will also be examined to see whether returns are systematically lower prior to bad unemployment news than good news. The tests are run on UK data from 1971 – 1993 and US data from 1951 – 1993. It is discovered that the evidence is difficult to reconcile with the types of incomplete market model that have been developed elsewhere in this thesis that give significantly different asset returns in the presence of nontradable risk to the complete market case. The problem arises because of the riskfree rate, which *rises* ahead of positive unemployment shocks. It is difficult to square this observation with precautionary savings models where the riskfree rate is endogenous. The *ex-post* equity premium is lower than usual prior to bad unemployment news which is consistent with Mankiw (1986) style models. It should be noted that the riskfree rate does fall (rise) *after* the poor (good) unemployment news. This is supportive of a finance model where investors react to, rather than anticipate, the announcement. However, the market index rises (falls) after a poor (good) unemployment announcement which is

evidence against a reaction (rather than anticipation) model. This chapter can be loosely interpreted as giving support for the economics type precautionary savings models where r_f is given over the finance models where r_f is inferred. Alternately, it can be reconciled with Heaton & Lucas' observation that asset returns are not driven by unemployment risk and that the results presented here are driven by other factors for which unemployment is just a proxy.

The chapter proceeds as follows. Section 28 develops the theory in a two period environment, discusses the implications of testing such a model on real data and examines existing literature on asset returns and unemployment shocks. Section 29 is the main empirical section of the chapter and tries to interpret the results of the tests run. Section 30 concludes. The *contribution* of this chapter is clear. This is the first test of incomplete market models that explains the Mehra & Prescott puzzles within a dynamic setting. That is, this is an empirical investigation of the first model to consider an economy where the individual risk to personal income changes over time and therefore investors do not know with certainty, in advance, the probability of receiving an endowment shock in the next time period. The results, particularly for the riskfree rate puzzle, are not encouraging. There are, though, several econometric and theoretical issues that emerge from this chapter that need resolving before Mankiw style models can be confidently rejected.

28 The economic environment

28.1 A one period model

Consider the following economy. At time 0 all investors are homogeneous and have consumption c_0 . At time $t = 1$ investors are either in the high income group with probability s and have consumption $c_1 = c_h$ or in the low income group and have consumption $c_1 = c_u \ll c_h$. s is not known with certainty at $t = 0$ and is assumed to have a probability density function $f(s)$. Assume that all dividend risk is linked to unemployment risk in keeping with the model of Mankiw (1986)¹⁰⁹ so that the dividend d^s is uniquely defined by the actual unemployment level s at $t = 1$. Because the price of the market and riskfree asset are known at $t = 0$, this means that equity premium ep^s over the interval $[0, 1]$ is also uniquely defined by the realisation of s . It will be shown that if $s > (<)E_0[s]$ then the observed excess return to the market over the riskfree rate must be lower (higher) than the time zero expectation over the interval $[0, 1]$. This is because, in order to have an *ex-ante* positive equity premium in this type of model, dividends must be lower in high unemployment states than low unemployment states. So, analysing *ex-post* returns should show a negative correlation between the equity premium and unemployment shocks. With regard to the riskfree rate, matters are a little more complicated. In the study to follow three month treasury bills are used as the riskfree asset and the time interval $[0, 1]$ is taken to be a year. Therefore, strictly speaking, at $t = 0$ there is not a riskfree asset that

¹⁰⁹To remove proper risk aversion considerations.

matures at $t = 1$. In order to invest risklessly for one period it is necessary to roll over the riskfree portfolio three times. This is, of course, not riskless as the price at which the three intermediate trades occur is not known with certainty at $t = 0$. What will be examined in the empirical section of this chapter is the change in the 3 month t-bill rate over the interval. If news comes into the market to causes investors to assign higher probabilities to high unemployment states at $t = 1$ then the precautionary savings motive should increase over the interval $[0, 1]$ and the rate of return offered on the riskless bond should drop. These relationships are demonstrated more formally below. Consider the one period Euler equation model. It is known that, for power utility and using $ep := r_m - r_f$:

$$E[ep] = \frac{-\text{Cov}(ep, c_1^{-\gamma})}{E[c_1^{-\gamma}]} \quad (58)$$

As the *ex-ante* equity premium is positive, so the right hand side of this equation is positive. This means that the covariance between the equity premium and $c_1^{-\gamma}$ must be negative. Let $f(s)$ denote the probability density function for possible unemployment rates:

$$\begin{aligned} \text{Cov}(ep, c_1^{-\gamma}) &= \int_0^1 (ep^s - E[ep])[(c_u^{-\gamma} - E(c_1^{-\gamma}))s + (c_h^{-\gamma} - E(c_1^{-\gamma}))(1 - s)]f(s)ds \\ &= [c_h^{-\gamma} - E(c_1^{-\gamma})] \int_0^1 (ep^s - E[ep])f(s)ds \\ &\quad + \int_0^1 s[c_u^{-\gamma} - c_h^{-\gamma}](ep^s - E[ep])f(s)ds \\ &= [c_u^{-\gamma} - c_h^{-\gamma}]E[s(ep - E[ep])] \\ &= [c_u^{-\gamma} - c_h^{-\gamma}]\text{Cov}(s, ep) \end{aligned}$$

Given that $c_h \gg c_u$ it is clear that $\text{Cov}(s, r_m - r_f) < 0$. The intuition is

straight forward. For a positive *ex-ante* equity premium the market portfolio must have a positive consumption beta. Therefore, the return on the equity premium portfolio must be positively correlated with consumption growth or negatively correlated with unexpected changes in unemployment. Alternately this can be interpreted as saying that, within a Mankiw style model, the positive *ex-ante* equity premium results from having lower dividends in states with a high risk of unemployment. Therefore, if we regress *ex-post* excess returns to the equity premium portfolio with unemployment shocks then we would expect to see a negative coefficient.

Next consider a riskfree bond which is created at $t = 0$ and matures at $t = 1$. Suppose that the bond can be traded at $t = -\delta t$ (before its creation) and at $t = 0$. At $t = -\delta t$, c_0 is known with certainty. At $t = -\delta t$ investors' probability density function for unemployment at $t = 1$ is given by $f(s)$. Between $t = -\delta t$ and $t = 0$, some news is revealed into the market that causes investors to change their beliefs so that their probability density function is now given by $f^*(s)$. Denote the price of the bond at $t = -\delta t$ and 0 by p_f, p_f^* respectively.

$$p_f^* - p_f = c_0^\gamma \int_0^1 [s c_u^{-\gamma} + (1-s) c_h^{-\gamma}] [f^*(s) - f(s)] ds$$

Any news that causes investors to place higher probabilities on high unemployment states give greater weight to the $c_u^{-\gamma}$ terms. As $c_h \gg c_u$, this increases the price of the riskfree bond. This implies that an increase in expectation of unemployment will reduce the riskless rate: this is a manifes-

tation of the precautionary savings motive described in chapter 3.

28.2 Application of model to real data

The theory developed above is in a one period world and is broadly in the style of Mankiw (1986). This predicts that the riskfree rate should decrease (increase) if investors' expectations of future unemployment increases (decreases) and that, ex-post, unemployment shocks and the excess return to the market index should be negatively correlated under rational expectations. This is, of course, an unusual phenomenon (although not unique) to observe. A reduction in the riskfree rate relates to a reduction in the discount rate and therefore, *ceteris paribus*, a rise in the market index.

Before proceeding, the problems in applying this single period model to multiperiod data is discussed. From the work of chapter 2 it is known that there are no theoretical problems in applying a single period Euler equation model in a multiperiod context. The problem is that c_1, c_0 are no longer clearly defined in the multiperiod case. First, consider c_0 . In the model described above there is *ex-ante* homogeneity of investors so that c_0 is the same for all. In a multiperiod environment people are, though, reentering as well as leaving the workforce and so c_0 should, ideally, not be fixed cross-sectionally. Second, the work of Heaton & Lucas has argued that c_h, c_l are not really a function of unemployment as financial markets allow investors to smooth consumption whatever their individual endowments. Therefore the seemingly innocuous assumption that $c_h \gg c_l$ is not necessarily true in this case. If a short term rise in unemployment can signal a long term decline in

the rate, then, under these conditions, a positive unemployment shock would be seen as good consumption news. In this case investors might consume more at times of positive unemployment shocks in anticipation of the future recovery. This issue is addressed in two ways. First, the first order autocorrelation for changes in unemployment are 86.4% and 17.6% for the UK and US respectively: both significant at the 0.1% level, implying that higher (lower) than expected unemployment at any point in time will lead to an increase (decrease) in investor's predictions of unemployment (and, indeed, change in unemployment) in the next time period. In this sense, positive unemployment shocks are permanent. Second, even if bad unemployment news is perceived as a signal of good future consumption news — and the positive autocorrelation in the rate of change of unemployment shocks suggests that this is unlikely — both the money and stock markets should interpret the news in the same way. The prediction that the change in riskfree rate and abnormal excess returns should have the same sign still holds.

It is also not clear within a real economy how the markets' expectations of unemployment changes over time. Clearly, at the time when the unemployment figure is announced, the market then has perfect knowledge of the shock. Within this study it is assumed that 12 months prior to the month to which the unemployment figure refers — which will be called month 0 throughout — the market estimates unemployment using an ARIMA model that is fitted in sample. Between month -12 and the time of announcement, the unemployment shock will reveal itself. If the markets anticipate this

news, then the change in riskfree rate and abnormal equity premium should have the same sign in the pre-event period. If the markets respond to the news, then the assumption about ex-ante homogeneity of investors no longer holds as the dismissals will already have taken place. However, the unemployment news will signal information about future unemployment changes. If the stock and money markets interpret this signal the same way, then the change in riskfree rate and abnormal equity premium should have the same sign in the post-event period. It should be emphasised that this in-sample method of calculating unemployment news is by no means the only one that was available to the author and is not necessarily the best either. A “one step ahead” forecasting technique or comparing real unemployment news against macroeconomic forecasts are mainstream alternatives. Analysing the sensitivity of the results presented against different forecasting techniques is a planned area for further research for the author.

Finally, within the “real word”, unemployment is clearly not only the variable that influences asset returns. The modelling process here will try to ensure that the observed changes in asset returns are not driven by changes in inflation. The empirical tests use nominal, not real, riskfree rates (for reasons explained below), and nominal rates are clearly influenced by inflation. Further, there are several studies — see for example Pindyck (1984) — that show the influence of inflation on stock prices. As mentioned below, inflation and unemployment shocks are not independent and so adjusting for inflation is clearly important. However, there are numerous other variables for which

unemployment might be proxying. Such examinations are considered outside the scope of this particular chapter. The rationale is that if unemployment is the key state variable determining the average equity premium and real risk-free rate then it should also be a key variable in determining intertemporal variations in asset returns.

From this discussion it is clear that, around the month of an unemployment shock, the change in the real riskfree rate and abnormal excess return to the market should have the same sign. If the market anticipates the shock, then the pattern should be observable in months $[-12, -1]$. If the market responds to the news, there should be no observable correlation between unemployment news and asset returns in period $[-12, -1]$ and the pattern should be in the post event period. Within a multiperiod context it is not possible to state categorically that positive (negative) unemployment shocks are related to negative (positive) changes in the riskfree rate/abnormal excess returns as unemployment news changes expectations about future consumption. Nevertheless, given the positive autocorrelation in changes in unemployment rate, it is most likely that the changes will be of this sign.

28.3 Unemployment shocks and asset returns

There have been three recent studies of the relationship between unemployment shocks and interest rates. Hardouvelis (1988) looks at the impact of 15 macroeconomic indicators on interest rate and exchange rate movements on the day of announcement. A negative correlation between unemployment changes and interest rate movements is discovered. Prag (1994) looks at

the change in interest rates with unexpected movements in unemployment on the day of an unemployment announcement in the US. It is found that, in general, there is a negative correlation between unemployment surprises and changes in interest rate. Below a certain unemployment level (the natural unemployment rate), the effect is amplified, probably due to inflationary worries. Bierens and Broersma (1993) look at the relationship between interest rate changes and unemployment. They argue that periods of high unemployment are often preceded by periods of high interest rates (the lag being 18 months in the US). They find a Granger causal effect in both the US and UK between interest rates and unemployment with unemployment being the dependent variable. This study therefore differs from the previous ones in two important respects. First, the relationship between interest rates and unemployment is positive. Second, in this case it is unemployment that is the dependent variable. Broadly speaking, this can be reconciled with the previous studies because of the 18 month lag between an interest rate high and an unemployment high. If we examine the *instantaneous* relationship between unemployment and interest rates at a time of high unemployment, interest rates will be falling, giving a negative coefficient despite the positive causal relationship. Few studies examining the impact of macroeconomic news on stock market data concentrate on unemployment. Papers examining cross-sectional variations in stock returns using the Arbitrage Pricing Theory with real economic factors (a stream of literature started by Chen et al. (1986) in the US and Poon and Taylor (1991) in the UK) do not use unemployment

as a variable. Similarly, Fama and French (1989) do not choose unemployment as a variable in trying to relate stock and bond returns to the business cycle. In a study of hourly stock market returns that does look directly at unemployment news, Jain (1988) finds that unemployment shocks have little influence. Sadeghi (1992), who examines Australian data, is the only paper, to this author's knowledge, that looks at longer term stock returns in the presence of unemployment news. He finds that stock market movements are positively correlated with simultaneous unexpected movements in the unemployment rate but are negatively correlated with revisions in the expected unemployment rate. However, while there is little work that directly examines the impact of unemployment shocks on equity returns, it may be that unemployment acts as a surrogate for other macroeconomic variables that have been examined. In particular, inflation changes and unemployment changes may be related. Prag (1994) argues that a rise in unemployment will lead to a decrease in expectations about inflation (and hence reduce the riskfree rate). This chapter will therefore attempt to separate the impact of inflation changes from those of unemployment shocks.

29 Empirical testing

29.1 Data

The intertemporal restrictions of the Mankiw (1986) style model are tested on US and UK data. The raw monthly data used in this study, together with the relevant source, is provided in table 7. Certain points need clarifying:

| Variable | Description | Source, Formula |
|-----------------------|---|--|
| Raw data | | |
| R_{ft} | Nominal monthly rfr | LSPD/CRiSP |
| R_{mt} | Monthly market return | LSPD/CRiSP |
| ep_t | Nominal monthly ep | $R_{mt} - R_{ft}$ |
| \bar{ep} | Average ep for sample | |
| U_t | Number unemployed, SA | Datastream |
| RPI_t | RPI — All items (UK) Consumer price index (US) | Datastream |
| I_t | Inflation, month t | $(RPI_t - RPI_{t-1})/RPI_{t-1}$ |
| Independent variables | | |
| $Shk\ 1_t$ | U_t ARIMA errors | See main text |
| S_t | Unemployment shock | $Shk\ 1_t - a - bI_t$ |
| Ω_t | Pre-event inflation | $\frac{RPI_{t-1} - RPI_{t-13}}{RPI_{t-13}}$ |
| Ω_t^P | Post-event inflation | $\frac{RPI_{t+6} - RPI_{t-1}}{RPI_{t-1}}$ |
| $\Delta I_{3,t}$ | Inflation change: pre-event | $\frac{RPI_{t+2} - RPI_{t-1}}{RPI_{t-1}} - \frac{RPI_{t-10} - RPI_{t-13}}{RPI_{t-13}}$ |
| $\Delta I_{3,t}^P$ | Inflation change: post-event | $\frac{RPI_{t+9} - RPI_{t+6}}{RPI_{t+6}} - \frac{RPI_{t+2} - RPI_{t-1}}{RPI_{t-1}}$ |
| Dependent variables | | |
| EP_t | Pre-event ep | $\sum_{i=-12}^{-1} ep_{t+i}$ |
| ΔR_{ft} | Pre-event change in rfr | $(1 + R_{ft-1})^{12} - (1 + R_{ft-13})^{12}$ |
| EP_t^P | Post-event ep | $\sum_{i=0}^{+6} ep_{t+i}$ |
| ΔR_{ft}^P | Post-event change in rfr | $(1 + R_{ft+6})^{12} - (1 + R_{ft-1})^{12}$ |

Table 7: Definitions of data used in this study. Figures for raw data not seasonally adjusted unless stated (SA = Seasonally adjusted). The riskfree rate is the return on the 90 day treasury bill, market returns include dividend reinvestment. The Consumer Price Index for the US is for all urban consumers, RPI=Retail Price Index. ep=equity premium and rfr=riskfree rate.

- Nominal, as opposed to real, returns are used throughout this study.

The reason for this is that month-by-month variability in the inflation rate is very much greater than in the nominal riskfree rate. Therefore, the correlation coefficient (in the UK) between inflation and the real riskfree rate over the sample period was -96.4%. Looking at the changes in the real rate on monthly data is essentially a surrogate for looking at minus inflation. The author therefore believes that the nominal rate gives a better indication of the precautionary savings motive. Using nominal rates requires that the method used is careful in its treatment of changes in inflation.

- The theory refers to the probability of becoming unemployed. Despite this, the empirical tests use the number unemployed (seasonally adjusted), as opposed to the unemployment rate, as the basis for calculating shocks. The reason for this is that the number unemployed each month was available to a greater number of significant figures. This should not give biased estimates of unemployment shocks. Further, given the number of redefinitions of unemployment in the UK, any “official” statistics cannot be said to give an unbiased estimate. This issue is not addressed further here.

Other raw data sources should be self explanatory. The independent variables used in the analysis to follow are also described in table 7. Again certain points need clarification:

- For the shock in unemployment, it was necessary to remove autocorrelation from the U_t series and so a Box-Jenkins approach was adopted. It was found that for both UK and US unemployment over the periods examined an ARIMA(1,2,1) with a seasonal lag 1 component on the AR and MA terms provided the “best” fit. This is an “unusual” ARIMA fitting but was used nonetheless as it was suggested by the data for both series.

The seasonal component was required on both series, which is perhaps surprising as the unemployment figures used were notionally seasonally adjusted already. Errors from these ARIMA models, denoted by $\text{Shk } 1_t$, had acceptable Box-Ljung autocorrelation statistics at all lags. The errors from this model, however, were correlated with inflation in the event month. As highlighted above, as nominal returns are being used, it is important to separate inflation effects from unemployment effects. Therefore, it was decided to orthogonalise $\text{Shk } 1_t$ from inflation in the event month using linear regression. That is, the unemployment shock used in the study was $S_t = \text{Shk } 1_t - a - bI_t$ where a, b are the regression coefficients from regressing $\text{Shk } 1_t$ against I_t in sample. Clearly now S_t is orthogonalised against instantaneous inflation although there remained some correlation between S_t and I_t at various lags. These correlations are, though, substantially lower than for the original $\text{Shk } 1_t$.

| | Ω_t | Ω_t^P | $\Delta I_{3,t}$ | $\Delta I_{3,t}^P$ |
|-----------------------|------------|--------------|------------------|--------------------|
| S_t (US) | -0.024 | -0.053 | 0.028 | -0.013 |
| S_t (UK) | 0.079 | 0.029 | 0.009 | -0.024 |
| Ω_t (US) | . | 0.786 | -0.144 | -0.141 |
| Ω_t (UK) | . | 0.634 | -0.124 | -0.110 |
| Ω_t^P (US) | . | . | 0.304 | -0.050 |
| Ω_t^P (UK) | . | . | 0.428 | -0.263 |
| $\Delta I_{3,t}$ (US) | . | . | . | -0.245 |
| $\Delta I_{3,t}$ (UK) | . | . | . | -0.320 |

Table 8: Correlation coefficients for independent variables used in linear regressions for the US and the UK.

- In order to further correct for inflationary effects, it was decided to include inflationary terms (Ω_t terms), which might be expected to influence abnormal equity premia and changes in inflation terms (ΔI_t terms), which might influence the change in the nominal riskfree rate.

The definitions of these variables are given in table 7.

29.2 Regression analysis

The main empirical tests consist of a series of linear regressions of the change in riskfree rate and the equity premium against unemployment shocks S_t . In total, the main regressions were run for 262 months for the UK (June 1971 – March 1993) and 514 months for the US (March 1951 – December 1993) using monthly data. The correlation coefficients between the independent variables are given in table 8. The dependent variables used in the analysis are also given in table 7. These are self explanatory. The main test of this chapter is conducted by running the following set of linear regressions:

$$\begin{aligned}
EP_t &= a + bS_t + c\Omega_t + d\Delta I_{3,t} + e_t \\
\Delta R_{ft} &= f + gS_t + h\Omega_t + j\Delta I_{3,t} + \epsilon_t \\
EP_t^P &= a^P + b^P S_t + c^P \Omega_t^P + d^P \Delta I_{3,t}^P + e_t^P \\
\Delta R_{ft}^P &= f^P + g^P S_t + h^P \Omega_t^P + j^P \Delta I_{3,t}^P + \epsilon_t^P
\end{aligned} \tag{59}$$

Inherent in these tests is the assumption that the expected equity premium in any month in the absence of inflation, inflation changes and unemployment shocks is constant over the sample. In this case, a and a^P will equal this average value. The theory developed above further predicts that $f = f^P = 0$. If not, there are factors affecting the rate of change of interest rate that neither inflation nor unemployment can obviously explain. It should also be noted that as Prag (1994) shows that the response of interest rates is not symmetric around the NAIRU, it would be desirable to split this regression to allow for this effect. This, again, is considered to be outside the scope of this chapter. In terms of relating the theory to the available data, the testable hypothesis is:

Hypothesis 1 *Either b, g should both be negative (markets anticipating shocks) or b, g should be insignificant and b^P, g^P should be negative (markets responding to shocks)*

There are clear problems here in that the dependent variables EP, EP^P are overlapping lag 11,6 respectively. The diagnostics on the basic OLS regressions (given in table 9) indicate serious heteroscedasticity and autocorrelation problems. In order to correct for these issues, all regressions (including those for $\Delta R_{ft}, \Delta R_{ft}^P$) are adjusted for the procedure of Newey and West (1987), Bartlett weights lag 12 (pre event) or 7 (post event). Table 9 outline

the results of the OLS regressions that were described in the equation set 59. While the diagnostics quoted are parametric, the Newey-West adjustments for heteroscedasticity and autocorrelation increase the robustness of these statistics.

In the pre-event period for both the UK and US the excess return to the market is negatively correlated with unemployment shocks to a high degree of statistical significance. This is in keeping with the theory developed in this chapter. However, the change in the riskfree rate is positively correlated with unemployment shocks, which is not consistent with the precautionary savings motive. Again, this effect is significant at very high confidence levels. In the post event period the riskfree rate is significantly negatively correlated with unemployment shocks, as the theory predicts if markets respond to unemployment news. However, the post event excess returns are positively correlated with unemployment shocks (significant at 5% in the UK, not significant at standard levels in the US).

| | a, a^P, f, f^P | b, b^P, g, g^P | c, c^P, h, h^P | d, d^P, j, j^P | Diagnostics |
|----------------|--------------------------|--------------------------|--------------------------|------------------------|---------------------------------------|
| EP | $1.45E^{-1}$ [4.89] | $-1.29E^{-4}$ [-3.38] | -1.72 [-2.32] | -1.23 [-0.77] | s.c.=412.9[0.000] het.=8.6[0.003] |
| EP | $9.75E^{-2}$ [1.49] | $-2.73E^{-3}$ [-3.75] | $-4.64E^{-1}$ [-0.56] | -5.46 [-2.59] | s.c.=198.0[0.000] het.=0.10[0.750] |
| ΔR_f | $-6.12E^{-3}$ [-2.34] | $1.33E^{-5}$ [3.23] | $1.60E^{-1}$ [3.06] | $8.61E^{-1}$ [3.99] | s.c.=376.1[0.000] het.=5.5[0.019] |
| ΔR_f | $-6.85E^{-3}$ [-1.00] | $1.83E^{-4}$ [2.39] | $8.41E^{-2}$ [1.16] | $6.98E^{-1}$ [2.79] | s.c.=207.8[0.000] het.=26.7[0.000] |
| EP^P | $8.87E^{-2}$ [4.87] | $3.64E^{-5}$ [1.13] | -1.58 [-2.70] | $6.81E^{-1}$ [0.74] | s.c.=374.9[0.000] het.=2.3[0.133] |
| EP^P | $3.87E^{-2}$ [0.77] | $1.09E^{-3}$ [1.95] | $-1.58E^{-1}$ [-0.14] | -1.72 [-1.84] | s.c.=195.1[0.000] het.=1.23[0.267] |
| ΔR_f^P | $-3.43E^{-3}$ [-2.04] | $-1.53E^{-5}$ [-3.58] | $1.19E^{-1}$ [2.27] | $6.26E^{-1}$ [3.56] | s.c.=344.0[0.000] het.=13.2[0.000] |
| ΔR_f^P | $-6.13E^{-3}$ [-1.40] | $-2.07E^{-4}$ [-2.72] | $1.16E^{-1}$ [1.45] | $3.40E^{-1}$ [2.06] | s.c.=196.7[0.000] het.=8.7[0.003] |

Table 9: An OLS regression of EP , EP^P , ΔR_f and ΔR_f^P against shocks in unemployment, inflation and changes in inflation. Figures in square parentheses associated with regression coefficients are t-statistics that have been Newey-West adjusted with Bartlett weights lag 12 or 7. The s.c. and het. diagnostics are LM test statistics (χ_1^2) for serial correlation and heteroscedasticity respectively, with associated p-statistics in square parentheses. Sample period: Jun 1971 – Mar 1993 (UK), Mar 1951 – Dec 1993 (US). The column headings should be read as follows: “ a, a^P, f, f^P ” is the constant in the regression, “ b, b^P, g, g^P ” is the coefficient of S_t , etcetera.

These results tie in with those of Sadeghi (1992) for the Australian market — the market is positively correlated with contemporaneous changes in unemployment but negatively correlated with future unemployment shocks. Further, the hypothesis that markets respond to unemployment shocks is not in keeping with the high levels of significance associated with $EP, \Delta R_f$, which should be insignificant in this case. b^P, g^P are also not of the same sign around the time of an unemployment shock. The author interprets these empirical findings as a rejection of the hypothesis that unemployment shocks can resolve the puzzles of Mehra and Prescott (1985).

29.3 Further empirical investigation

In order to study the effect in more detail, an “event study” was run looking at the changes in the equity premium and riskfree rate around the time of the shock in unemployment. The sample periods for the two markets were divided into equally sized (up to rounding error) quartiles depending on whether the month had a “very positive”, “positive”, “negative” or “very negative” shock in unemployment as measured by S_t . Abnormal returns to both the equity premium and changes in the riskfree rate were then calculated in the months $[-12, +6]$ relative to the event month (the month to which the unemployment figure refers) and cumulated. Here the cumulative abnormal returns for the equity premium p_{it}^t and cumulative monthly changes in real riskless rate f_{it}^t are defined as (the m superscripts denote that it is monthly returns that are being used — although by taking the riskfree rate terms to the power of 12 these are being annualized):

$$f_{ij}^t := \sum_{k=t+i}^{t+j} (1 + R_{fk}^m)^{12} - (1 + R_{fk-1}^m)^{12}$$

$$p_{ij}^t := \sum_{k=t+i}^{t+j} ep_k^m - \bar{ep}^m$$

Notice that if $i = -12, j = -1$, then $f_{ij}^t = \Delta R_{ft}$ and $p_{ij}^t = EP_t - \bar{ep}^{12m}$ (\bar{ep}^{12m} is the average annual equity premium). Similarly, if $i = 0, j = +6$, then $f_{ij}^t = \Delta R_{ft}^P$ and $p_{ij}^t = EP_t^P - \bar{ep}^{7m}$ (\bar{ep}^{7m} is the average seven month equity premium). Let f_{ij}, p_{ij} be used to denote the average values of f_{ij}^t, p_{ij}^t over the sample periods, with an associated standard deviation. This is, of course, not a standard event study as there are 19 events (unemployment shocks) over this period. However, as unemployment shocks are random, the 18 “events” in the months other than month 0 should “cancel out”. Figures 13 and 14 show the event cumulative abnormal returns for the equity premium and changes in the riskfree rate against unemployment shocks in month 0 for $i = -12, j = +6$ in the UK — the graphs for the US are not presented here. As can be seen, there is a significant pattern in both the riskfree rate and the equity premium in the run-up to an unemployment shock, suggesting that the markets do indeed anticipate the shock prior to the event month. In response to the news, both series appear to “revert”.

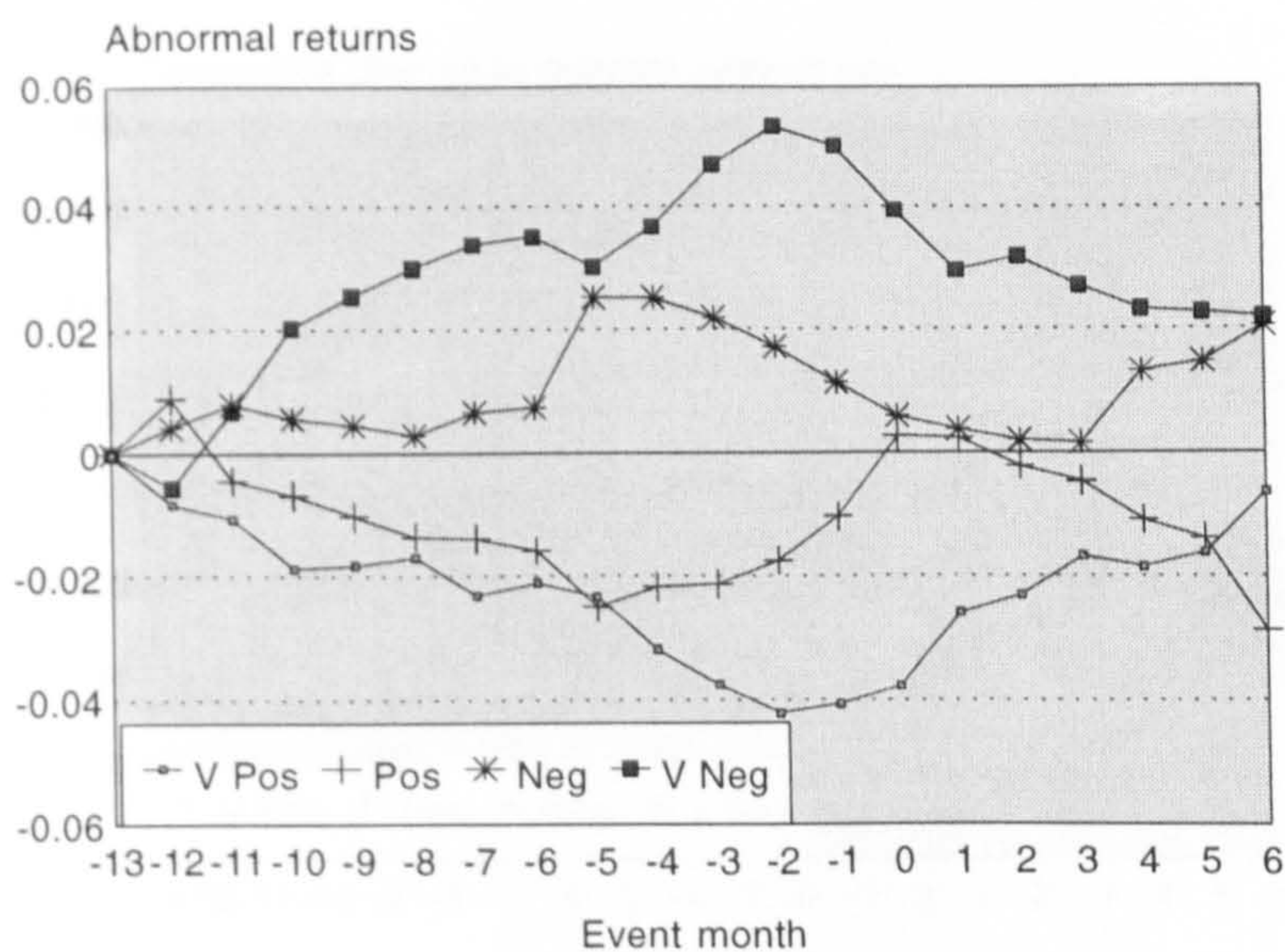


Figure 13: Cumulative abnormal returns to the equity premium around a shock in unemployment as modelled by S_t . Here “abnormal return” at month t is the difference between the equity premium in that month and its average over the whole sample: p_{ij} , $i = -12, j \in [-12 + 6]$. Unemployment shocks are divided into equally sized quartiles. Month 0 is the month referred to in the unemployment statistic. The sample period is June 1971 to March 1993 and is based on UK data.

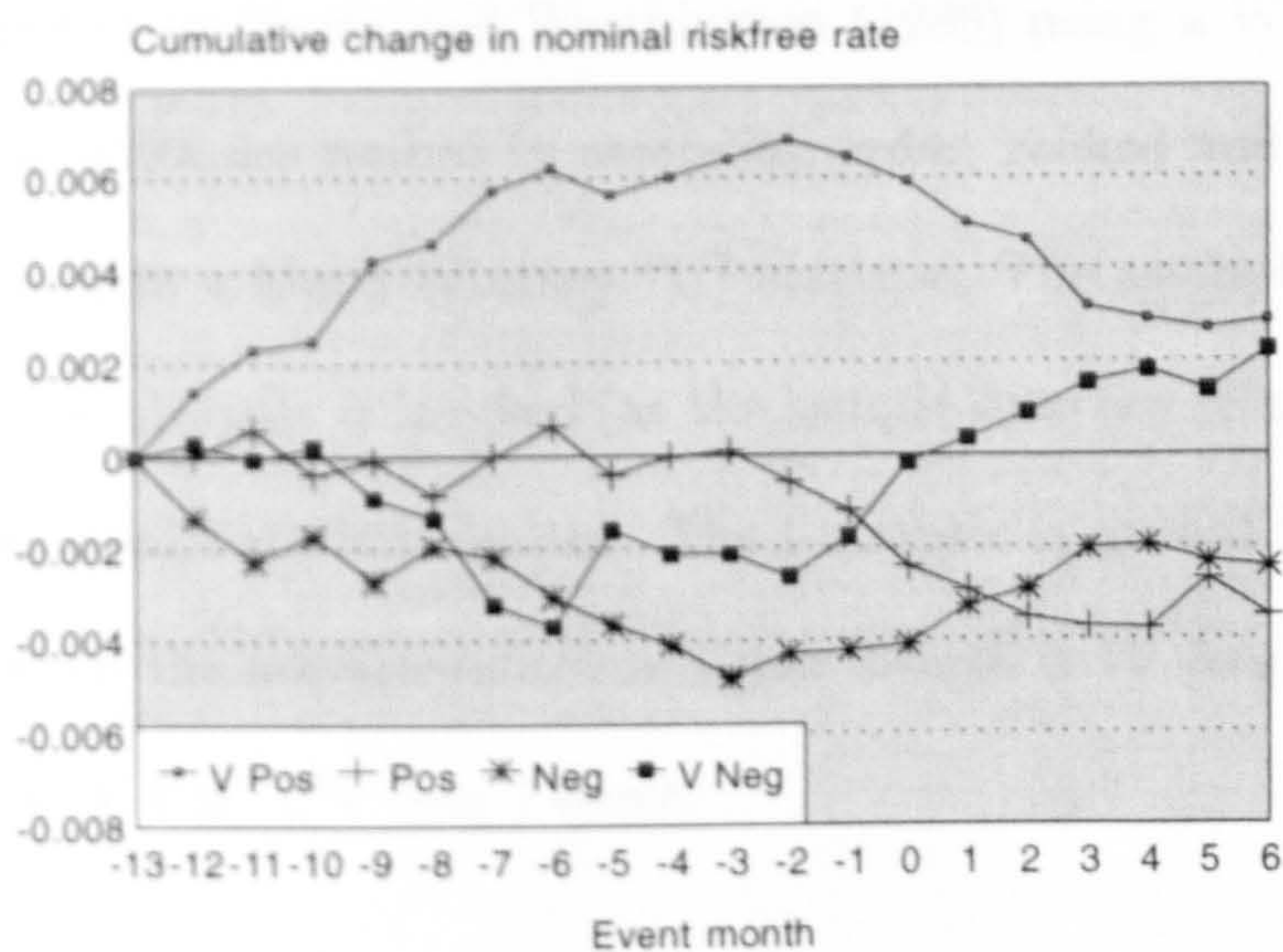


Figure 14: Cumulative changes in the nominal riskfree rate around a shock in unemployment as modelled by S_t : f_{ij} , $i = -12, j \in [-12, +6]$. 0 on the y-axis refers to the riskfree rate in month -13. Unemployment shocks are divided into equally sized quartiles. Month 0 is the month referred to in the unemployment statistic. The sample period is June 1971 to March 1993 and is based on UK data.

Table 10 examines the magnitude of this effect for f_{ij}, p_{ij} with $i = -12, j = -1$ and $i = 0, j = +6$ in both the US and UK. That is, table 10 gives the average values (with associated standard deviations) of $EP_t, \Delta R_{ft}, EP_t^P$ and ΔR_{ft}^P around “V Pos”, “Pos”, “Neg” and “V Neg” unemployment shocks in the sample periods. Table 11 gives a non-parametric test of the significance of these results around extreme unemployment shocks. The technique used is that suggested by Neave and Worthington (1988) using a Wilcoxon sign test. The variables are ranked in ascending order, ranked and the sum of ranks converted to a Mann-Whitney “U” statistic. The asymptotic normal property of this statistic is invoked (as the sample sizes are relatively large) and the associated t-statistic quoted. The t-statistic is quoted for “V Neg” (“V Pos”) when the average return is higher around a “V Neg” (“V Pos”) shock than a “V Pos” (“V Neg”) shock.

| | f_{ij} | | p_{ij} | |
|-------------|----------------------|----------------------|---------------------|---------------------|
| | $i = -12, j = -1$ | $i = 0, j = +6$ | $i = -12, j = -1$ | $i = 0, j = +6$ |
| V Pos US | 0.00224 (0.0173) | -0.00264 (0.0137) | -0.02002 (0.163) | 0.02114 (0.120) |
| V Pos UK | 0.00715 (0.0253) | -0.00360 (0.0184) | -0.0438 (0.228) | 0.0356 (0.163) |
| Pos US | 0.00160 (0.0162) | 0.00031 (0.0112) | 0.00525 (0.147) | -0.01386 (0.111) |
| Pos UK | 0.00142 (0.0237) | -0.00201 (0.0222) | -0.0159 (0.226) | -0.0203 (0.204) |
| Neg US | -0.00044 (0.0181) | 0.00494 (0.0174) | 0.00514 (0.148) | -0.00865 (0.113) |
| Neg UK | -0.00536 (0.0280) | 0.001480 (0.0218) | 0.0109 (0.254) | 0.0118 (0.187) |
| V Neg US | -0.00180 (0.0189) | 0.00296 (0.0145) | 0.00963 (0.148) | 0.00136 (0.123) |
| V Neg UK | -0.00088 (0.0283) | 0.00419 (0.0197) | 0.0488 (0.243) | -0.0270 (0.173) |

Table 10: Average change in nominal riskless rate f_{ij} and cumulative abnormal equity premium p_{ij} around the time of an unemployment shock as modelled by S_t . Standard deviations (not errors) are given in parentheses. Sample period: Jun 1971 – Mar 1993 (UK), Mar 1951 – Dec 1993 (US)

As can be seen, the excess return to the market is significantly higher prior to a very negative unemployment shock than a very positive unemployment shock for both the US and UK as the theory outlined above would predict. However, the change in the riskfree rate is more positive prior to a very positive shock than a very negative shock. This is significant at the 5% level in the US (but not quite in the UK) and is not consistent with the theory. Conversely, in the post event period, the riskfree rate moves in the predicted direction but the excess returns do not. It is difficult to reconcile tables 10, 11 with the theory outlined above.

It has been argued in the preceding section that the results obtained are not consistent with a model in the style of Mankiw (1986) for explaining the Mehra & Prescott puzzles. Can the data be explained by any alternate model? Consider the conclusions of Bierens and Broersma (1993). They argue that changes in the riskfree rate is the independent variable and that the change in unemployment is driven by changes in this rate. So suppose that changes in the riskless rate are exogenous. Under the Gordon Growth Model, for given risk premium r_p and growth rate g , then (see, for example,

| | EP | EP^P | ΔR_f | ΔR_f^P |
|------------|------|--------|--------------|----------------|
| US - V Pos | . | 1.37 | 2.27 | . |
| US - V Neg | 1.86 | . | . | 3.15 |
| UK - V Pos | . | 2.55 | 1.69 | . |
| UK - V Neg | 2.50 | . | . | 2.26 |

Table 11: The t-statistics associated with the Wilcoxon non-parametric test for between difference in means between "V Pos" and "V Neg" for pre and post event abnormal equity premiums and changes in riskfree rate. The t-statistic is quoted for the sample with the higher average values.

p. 243 of Franks, Broyles and Carleton (1985)):

$$p_m = \frac{D}{r_f + r_p - g}$$

So, if changes in the riskfree rate do not influence g, r_p, D , then the instantaneous response on the market r_m^i to a change in the riskless rate from r_f to r_f^* is:

$$r_m^i = \frac{r_f - r_f^*}{g - r_f^* - r_p}$$

On the assumption that $g - r_f^* - r_p = 5\%$ ¹¹⁰, this predicts that $EP = -\Delta R_f/0.05$ and $EP^P = -\Delta R_f^P/0.05$. This proposition was tested using a standard t-test for all 262 months for the UK and 514 months for the US. The t statistics were 0.38 and 0.52 in the pre-event period and 0.01 and 0.43 for the post-event period for the UK and US respectively. This, then, appears supportive of a model where the riskfree rate is exogenous and the stock market and unemployment react to this variable.

30 Conclusion

This chapter was motivated by the substantial existing research that examines whether the Mehra & Prescott puzzles can be explained by disaggregating consumption to allow for unemployment risk. An empirical implication of these models is that shocks in unemployment should result in substantial movements in the riskfree rate and the equity premium. This is the first piece

¹¹⁰The average dividend yield over the period in the UK is 5.02%, with a range of 2.85% to 11.77% — source: Datastream.

of research to consider an economy where investors do not know in advance exactly what their risk of suffering an endowment shock in the next period is. This risk also changes from period to period. If unemployment turns out to be higher than expected then the market's return will be lower than expected over the interval as dividends shocks coincide with income shocks in a style reminiscent of the model of Mankiw (1986). Further, as the market becomes more aware that the risk to personal capital is higher than was previously imagined, so the precautionary savings motive rises and the riskfree rate drops. So, it is expected that returns to the market will drop at a time when the riskfree rate is also falling around the time of poor unemployment news. By an analogous argument, if the risk to personal capital is lower than expected then the stock market and returns to treasury bills should both rise.

As outlined in the previous section, the data does not appear to be consistent with these models. Instead, it appears that the riskfree rate is the independent variable, with returns to the stock market responding to this. This is because the riskfree rate rises (falls) prior to poor (good) unemployment news. While returns to treasury bills do drop after the announcement, the stockmarket rises during this period. The main conclusion of this chapter is that unemployment risk is unlikely to be the explanation of the Mehra & Prescott puzzles in a Mankiw-style model. This, though, is not to say that using disaggregated consumption will not explain the puzzles, just that using unemployment would not appear to be the most suitable way of doing

the disaggregation. Examining alternate ways of breaking down aggregate consumption data that will help explain both average asset returns and short term movements in these returns is an area that requires further examination.

Part VII

Conclusion

31 Conclusion

This thesis is concluded with three subsections. First, a summary of findings is presented. Then several areas for future research are discussed. The thesis is completed with some final observations.

31.1 Summary of findings

This thesis has analyzed the ability of incomplete market models to explain two well documented financial market anomalies — the “Mehra & Prescott puzzles” (riskfree rate puzzle and equity premium puzzle) and the dividend controversy. In chapter 2, a solid theoretical foundation for the thesis was reviewed, starting with the most general asset pricing framework of all — the fundamental theorem of asset pricing. Power utility functions were then discussed. That the parameter γ plays two roles in this utility function, describing both the instantaneous risk aversion and elasticity of intertemporal substitution, was emphasised. Then, within an equilibrium setting, it was shown that the current price of any asset is given by the expectation of its payoff in the next period multiplied by the ratio of marginal utilities of next period’s consumption to this period’s consumption. This is the Euler equation which lies at the heart of this thesis. The chapter then considers environments where a representative agent exists. It was stressed that only with complete markets can aggregate consumption be used in the Euler equation to determine equilibrium asset prices. If the market is incomplete, as assumed throughout this thesis, then individual consumption, which is more

volatile than the aggregate, should be used instead. Next, the chapter considered both discrete and continuous time developments of the consumption CAPM. Essentially, provided that consumption of all investors is smooth and the utility function is “sensible”, then the CCAPM follows. But if individual consumption and asset returns are not smooth and the market is incomplete then not only will the CCAPM not follow, but violations from the model will depend on the allocation of idiosyncratic risk. Finally, chapter 2 uses four models (single period non parametric, single period parametric, multiperiod discrete state and a Merton (1971) style analysis) to provide expressions for the equity premium and real riskfree rate. These provide the basis for analysis in chapter 3.

Having provided a comprehensive theoretical basis for development, chapter 3 is probably the most extensive review of the Mehra & Prescott puzzles that currently exists in the literature. It was shown that, under representative agent assumptions, the puzzles can be described using any of the four formulations for the equity premium and real riskfree rate given in chapter 2. The intuition is clear; aggregate consumption is growing so fast and so smoothly it is not clear why the demand for borrowing does not outstrip the supply of savings given the low average real riskfree rate. Interest rates, apparently, needed to be higher on average to curb the incentive for borrowing and encourage saving in the US over the past century. That aggregate consumption has been so smooth means that it is difficult within a CCAPM framework to generate a large risk premium on *any* asset. What became

clear from this analysis is that the puzzles arise from a combination of four assumptions: (i) a representative agent exists (ii) aggregate consumption is smooth (iii) utility takes power form with $\beta < 1, \gamma < c.6$ and (iv) there are no taxes or market frictions. It is argued that there are three potential explanations for the Mehra & Prescott puzzles. First, it was argued that the *ex-post* observed average excess return to the market and real riskfree rate does not necessarily represent the *ex-ante* expectation. Second, certain forms of investor preferences were examined that separate the coefficient of relative risk aversion from the elasticity of intertemporal substitution. These are, in general, unable to resolve the equity premium puzzle. Third, we concentrated on incomplete market explanations of the puzzles both in the presence and absence of market frictions. The economics literature on precautionary saving, where the riskfree rate is exogenous and consumption endogenous, was briefly reviewed. While the findings of this literature are not wholly unambiguous, the balance of evidence presented does point toward the existence of substantial precautionary saving. We then concentrated on models with exogenous aggregate consumption but where asset returns are endogenous. It was argued that the persistence of idiosyncratic risk is the key factor determining the ability of incomplete market models to explain Mehra & Prescott's puzzles. Given that personal income risk seems fairly short term, particularly in the United States, only with severe market frictions can the puzzles be resolved. Chapter 3 was concluded with a very brief review of empirical tests of the CCAPM. This provided estimates of the coefficients

β, γ used elsewhere in the thesis and showed that the Mehra & Prescott puzzles are just one of a number of anomalies that arise from consumption based asset pricing models.

Chapter 4 took one key paper on incomplete market explanations of the Mehra & Prescott puzzles, Weil (1992a), to examine how sensitive the quantitative implications of incomplete market models are to the precise functional form of the utility function. Unlike existing studies examining investor preferences in this context, additive time separability is kept in this chapter. Instead, it was argued that there is a whole class of utility functions that “look” similar close to the point of expected future consumption are very different at points of low endowment. Indeed, it was shown that it is virtually impossible, given our current understanding of investor preferences, to accurately determine the magnitude of any incomplete market effect in this case. This chapter also provided a unifying approach to local proper risk aversion. From this analysis, a new form of proper risk aversion emerged naturally (labelled “Core Basic Risk Aversion”), which locally has the same necessary and sufficient conditions as the less general local Standard Risk Aversion.

Perhaps it can be said with some confidence at this stage, that income shocks that are independent of marketable jumps will not, on their own, be a true explanation for the equity premium puzzle. That is, proper risk aversion is not powerful enough to be a full explanation for Mehra & Prescott’s puzzles on its own. Within Weil’s example, dividends can account for up to 50% of

consumption in low endowment states (or, indeed, 100% of consumption if unemployment income is zero). The required contribution of dividends to overall consumption seems too high to be realistic. I do not wish to understate the importance of the proper risk aversion literature; instead the conclusion is that its application to this particular anomaly may be limited. This conclusion is strengthened by the results of chapter 4 which show that proper risk aversion, on its own, does not produce equity premia of the magnitude reported by Weil.

Chapter 5 considered the application of incomplete market models to the dividend controversy. It might, perhaps, be best understood as a variation in the model of Jensen (1986) where there is a systematic tendency for firms to overinvest due to agency problems. Dividends are paid in this case to try to minimise this overinvestment. In the economy of chapter 5, though, there are no conflicts between managers and investors. The main assumption that drives the chapter is that investors are uncertain about the aggregate level of future investment and cannot be sure that this investment will be optimal. Two effects combine to provide the optimal aggregate dividend policy. First, investors want *positive* investment shocks concentrated into states with low risk to personal capital. Second, simulation results demonstrate that even mean zero investment “surprises” should be concentrated in states where the risk to individual endowment is low. This was demonstrated in a two period model where financial assets can be traded and in a multiperiod model where trading is prohibited. It is argued that investors can reduce the risk

of overinvestment in states of high risk to nontraded assets by ensuring that, at the portfolio level, rights issues are concentrated into bull markets and dividends are only cut as a measure of last resort in bear markets. While there is no existing empirical studies that gives direct evidence on the validity of this model, it is argued that the findings of previous work does broadly supports the simulations in this chapter.

Chapter 6 argued that, if the risk of uninsurable personal income shocks is driving the *average* returns to stocks and treasury bills then the change in this risk should influence asset returns. Increased precautionary savings motive means that the riskfree rate should fall prior to high endowment risk. That the *ex-ante* equity premium is positive implies that market risk should be positively correlated with consumption risk: that is, the *ex-post* return to the market should be lower than average when there is a positive realised unemployment shock. This hypothesis was tested on both UK and US data. It was discovered that the market does indeed underperform in the twelve months leading up to bad unemployment news. However, the riskfree rate rises during this interval, supporting the findings of Bierens and Broersma (1993). This is extremely difficult to reconcile with finance-style precautionary savings models where the riskfree rate is endogenous. Returns to treasury bills do start to fall after a positive unemployment shock, but as stock market returns are higher than average over this post-event period, it is difficult to see how these two observations could be reconciled by incomplete market models. It was concluded that this observation is consistent with the

hypothesis that unemployment shocks are too short-lived to have a significant impact on equilibrium asset returns.

In subsection 31.3 I will draw some general conclusions from the different studies described in the body of this dissertation. Incomplete market models are, though, in their infancy and it not possible at this stage to make many inferences with great certainty. As emphasised in chapter 1, many questions addressing the ability of incomplete market models to resolve financial market anomalies remain unanswered. In the subsection below, an outline is presented of some of the future research that needs to be undertaken before we can say with any certainty whether the Mehra & Prescott puzzles and corporate finance anomalies can be resolved by introducing uninsurable risk into asset pricing models.

31.2 Further research

Many additional tests are needed to determine the ability of incomplete market models to explain financial market anomalies that will keep theorists and empiricists busy for many years to come as it is only in the last few years that multiperiod incomplete market models have started to appear in the consumption based asset pricing literature. Much further work is needed. In the opinion of the author, theorists will turn increasingly to computer generated simulations to determine the implications of their models. Empirically very little is known about the way stocks returns, in particular, are influenced by income shocks. Because of the potential for further research in this area, this subsection can only highlight a few of the paths for further investigation

that may prove fruitful. This subsection is divided into two parts. First, the most direct extensions to the three main chapters of substantive original contribution in this thesis are discussed. I then briefly consider some other, more general, extensions to the existing incomplete market literature that would appear to be a worthwhile area for future research.

- The key result from chapter 4 concerns the functional form of the utility function that we assume is shared by all investors. Several utility functions that are very similar close to the point of expected future consumption are extremely different in low consumption states. Existing literature really provides no strong support for power utility over other additively time-separable utility functions. Its popularity appears to come from a combination of tractability and the fact that to assume constant relative risk aversion is not blatantly counterfactual. As, to this author's knowledge, there has been no specific test of the risk aversion of individuals in low endowment states, the continued use of power utility in incomplete market models should not continue unquestioned. Without necessarily relaxing the assumption of fixed preferences across investors, it is of great importance that work is undertaken to specifically examine the risk aversion of agents in low income states. Chapter 4 also provides an integrated approach to local proper risk aversion. This literature does, though, remain reasonably abstract and work on the applicability of the theorems to "real world" problems is an area that needs further investigation.

- Chapter 5 argues that the possibility of low endowment states should influence aggregate dividend policy as the total demand for dividends will be high when there is a high probability of low income. There are several issues that this chapter raises that require further examination. Chapter 5 is based on simulation and further research is needed into both the theoretical and empirical consequences of this type of economy. From a theoretical standpoint, a joint model where companies issue bonds as well as equity might provide an insight into optimal capital structure issues as well as optimal dividend policy. This is because the timing of cashflows is secure with bonds and so, default excluded, there is no equivalent of dividend cut / rights issue in this case. Ideally trading should be introduced into the multiperiod simulations to allow for consumption smoothing through the buying and selling of financial assets, although given the number of potential paths through a many period environment, creating such a model will be a major undertaking. The main empirical implication of the model is that rights issues and dividend changes will be received more favourably at times of low risk to personal capital than periods of high risk. This is, in principle, straightforward to test using standard event study methods. The author is surprised that more tests of the reaction of shareprices to dividend changes have not concentrated on intertemporal (as opposed to cross-sectional) issues. This would certainly seem to be a useful route for further study. However, separating out the effect of changes in per-

sonal risks on share price reaction to dividend changes from the effects of other procyclical market phenomena might prove difficult. Further, as emphasised in chapter 6, we cannot yet identify the main sources of personal endowment risk.

- Chapter 6 considers how asset prices move as unemployment risk changes.

The results indicate that the riskfree rate rises prior to poor unemployment news, which is difficult to reconcile with a finance-style precautionary savings model where the riskfree rate is endogenous. While the conclusions seems robust to several statistical tests, both parametric and non-parametric, there are certain econometric issues that would benefit from further investigation. In particular, different definitions of unemployment shock might be used (for example, one step ahead forecasting) to see how sensitive the results are to the ARIMA model used. Further, this problem might be usefully tackled using a vector autoregressive approach.

Theoretically, it would be desirable to develop a model within a multi-period or continuous time framework where, rather than there always being a fixed probability s of unemployment next period, there is a time varying probability density function $f_t(s)$ of low endowment. Indeed, it is easy to envisage an “unemployment CAPM” emerging from this type of environment. At the most superficial level, this might be considered to be a combination of the Mankiw-style model developed in chapter 6 and the jump-diffusion CCAPM model of Back (1991) de-

scribed in chapter 2. Cross-sectional variations in asset returns could also be examined in this context.

These are the most direct areas for further research that arise from the three main studies in this thesis. There are many other more general areas of research into incomplete market models that might prove extremely fruitful, a few of which are now discussed:

- The question arises as to what constitutes endowment risk. Throughout this thesis, and particularly in chapter 6, the link between unemployment and income uncertainty has been made. It is by no means certain, though, that this is the risk to which incomplete market models are most easily applied. It may well emerge that early retirement is a more important source of uninsurable risk in determining equilibrium asset prices than unemployment. The author is surprised that, with the exception of some precautionary savings models in the economics literature, life-cycle model have not been more prevalent in existing incomplete market studies. An overlapping generations model with stochastic times of retirement and death would appear to be a sensible path for future investigation. This has two major advantages over the unemployment models presented in this thesis. First, retirement is persistent and so one of the main weaknesses of multiperiod unemployment models is overcome. Second, as much individual saving occurs later in life, it is perhaps easier to see retirement influencing savings decisions than unemployment risk.

- Throughout this thesis, as is standard with incomplete market models, it is assumed that investors save directly. Short-term idiosyncratic risk endowment can be smoothed in a Heaton & Lucas style model through the purchase and sale of financial assets. However, on the London Stock Exchange, only around one fifth of the total value of equity is owned directly by individual investors. Pension funds and insurance companies are the most important groups of shareholder. It may well be that the process of financial intermediation has a significant effect on the implications of incomplete market models. In particular, it is difficult for an investor to sell pension fund holdings in order to short-term consumption smooth. How pension funds react to periods of high unemployment (low pension fund contribution) would be an interesting area for further research. It would also be interesting to see the effect of *aggregate*, Rietz (1988) style, shocks on insurance companies' portfolio holdings.
- As highlighted throughout this thesis, it is the persistence of personal endowment risk that determines the effect of incomplete market assumptions on predicted equilibrium asset prices. There seems to be no consensus in the literature on this point. In chapter 3 it was shown that severe income risk, particularly in the US, appears to be fairly short-lived. On the other hand, chapter 2 shows that consumption growth does vary considerably between investors. So, existing work does not provide accurate estimates of the actual level of persistence

facing individual investors. More detailed analysis of the cross-sectional distribution of income across equity holders is vital if we are to determine whether the models of Heaton & Lucas or Mankiw are more likely to describe the real economy.

- It should be emphasised why it has been so difficult to explain the average observed returns to stocks and treasury bills over the last century at the same time as explaining the volatility of these returns. In order to have a high equity premium and high volatility in the equity premium, the pricing kernel should also be highly volatile¹¹¹. However a volatile pricing kernel should imply a variable riskfree rate. This is at odds with the smooth time-series that is observed for real treasury-bill returns. One way of immediately resolving the riskfree rate puzzle would be to assume that spot interest rates are an exogenous variable to finance models and are instead set centrally and determined by broader macroeconomic considerations. This possibility was briefly discussed in chapter 6. An incomplete market model with the riskfree rate predetermined but where the expected return and volatility to the stock market index and (say) a well diversified bond portfolio are endogenous would be an interesting and innovative route to pursue.

This list for potential future research into incomplete market explanations for financial market anomalies is certainly not exhaustive. Hopefully,

¹¹¹Notice that this discussion can be expressed in terms of the fundamental theorem of asset pricing itself rather than the more specific Euler equation.

though, it will motivate the reader to believe that this is an area currently under-researched and where there is great potential for important new discoveries. On a more personal note, the author's own immediate plans for future research concentrate on developing a multiperiod simulation of the market described in chapter 6 where there is uncertainty about the probability of unemployment in the next period and where unemployment risk varies from period to period. The dynamics of asset returns in this context can hopefully then be better understood¹¹².

31.3 Final comments

Given that such a large percentage of wealth comes from sources other than equity (dividends currently comprise around 4% of GNP in the UK), to assume that all assets can be readily traded does not seem realistic. The empirical evidence on the cross-section of individual consumption presented in chapter 2 shows that to assume that everyone is fully insured for all potential outcomes is not reasonable. The results of chapter 5 also suggest that the incomplete market paradigm might be applied to a wider range of puzzles than had previously been considered. To this author, the question is not, then, *whether* shocks to personal capital should be incorporated in asset pricing models (and also, potentially, theorems of optimal corporate finance policy), but how this task should be undertaken.

¹¹²In the period immediately following the submission of this thesis, the author is to visit Northwestern University on sabbatical. As the institute of Heaton & Lucas, where the "auctioneer's algorithm" has been developed, this is arguably the leading centre for multiperiod incomplete market simulations at the time of writing.

Detractors of incomplete market models may argue that the results of Heaton & Lucas, as well as the findings of chapter 6, give little support to this paradigm. This author does not agree with these sentiments. First, while Heaton & Lucas show the theoretical importance of persistence in personal risk, their cross-sectional evidence on income shocks is not sufficiently detailed to show that no important sources of persistent endowment risk exist. Further, the cross-sectional evidence on consumption does imply strongly that investors are not able to fully smooth personal risks. Chapter 6 focuses on unemployment as the source of risk and therefore is not a general test of incomplete market models. Also, the results of chapter 6 are not necessarily inconsistent with a model driven by unemployment if, in the style of an economics model, the riskfree rate is made exogenous.

It should be noted that, if an apparent inability to fully resolve Mehra & Prescott's puzzles is seen as sufficient grounds for rejecting incomplete market models, then we must also reject whole classes of complete market model as well. Despite the vast amount of research time that has been spent by many authors trying to settle the anomalies, there is still no generally accepted solution. There is a fundamental problem trying to reconcile a high average, and highly volatile, equity premium with a smooth real riskfree rate. The only "simple" explanation for the puzzles is that the long term first and second moments of consumption growth and asset returns in the US and UK are not representative of the *ex-ante* expectations. Indeed it is tempting to infer that the equity premium figure that is usually placed in the CAPM is

positively biased as the standard textbook treatment is to use a long term historic average value.

In conclusion it is argued that both the extensive literature review presented in chapters 2 & 3 and the three main chapters of substantive original contribution have shown that uninsurable risk may play an important role in resolving financial market anomalies. While the results have been mixed — chapter 5 has demonstrated an important new application of the theory to aggregate dividend policy while the results of chapter 6 are difficult to reconcile with a standard finance-style (riskfree rate endogenous) precautionary savings model — the balance of evidence suggests that this is an important area for further research. Important discoveries are likely to emerge as the most significant sources of individual endowment shocks are identified.

Part VIII

Bibliography

References

- Aase, K. A.: 1993, A jump/diffusion consumption-based capital asset pricing model and the equity premium puzzle, *Mathematical Finance* **3**, 65 – 84.
- Aase, K. A.: 1996, An equilibrium approach to derivative securities: stochastic volatility and survival, Mimeo: Norwegian school of economics and business administration.
- Abel, A. B.: 1990, Asset prices under habit formation and catching up with the Joneses, *American Economic Review* **80**, 38–42.
- Abel, A. B.: 1994, Exact solutions for expected rates of return under Markov regime switching: implications for the equity premium puzzle, *Journal of Money, Credit & Banking* **26**, 345–61.
- Ahn, C. M.: 1989, The effect of temporal risk aversion on optimal consumption, the equity premium and the equilibrium interest rate, *Journal of Finance* **44**, 1411–20.
- Ahn, C. M.: 1990, A resolution of the equity premium puzzle, *Economics Letters* **32**, 153–6.

- Aiyagari, S. R.: 1994, Uninsured idiosyncratic risk and aggregate saving, *Quarterly Journal of Economics* **109**, 659–84.
- Aiyagari, S. R. and Gertler, M.: 1991, Asset returns with transactions costs and uninsured individual risk, *Journal of Monetary Economics* **27**, 311–31.
- Ang, J. S.: 1987, *Do dividends matter? A review of corporate dividend theories and evidence*, Monograph Series in Finance and Economics: Salomon Brother Center for the Study of Financial Institutions.
- Arrow, K.: 1970, *Essays in the theory of risk-bearing*, Amsterdam: North Holland.
- Attanasio, O. P. and Davis, S. J.: 1996, Relative wage movements and the distribution of consumption, *Journal of Political Economy* **104**, 1227–62.
- Back, K.: 1991, Asset pricing for general processes, *Journal of Mathematical Economics* **20**, 371–95.
- Bansal, R. and Coleman, W. J.: 1996, A monetary explanation of the equity premium, term premium, and risk-free rate puzzles, *Journal of Political Economy* **104**, 1135–71.
- Bar-Yosef, S. and Huffman, L.: 1986, The information content of dividends: a signalling approach, *Journal of Financial & Quantitative Analysis* **21**, 47–58.

- Benninga, S. and Protopapadakis, A.: 1990, Leverage, time preference and the equity premium puzzle, *Journal of Monetary Economics* **25**, 49–58.
- Bentzen, E. and Sellin, P.: 1997, The Intertemporal Capital Asset Pricing Model with returns that follow Poisson jump-diffusion processes, Mimeo: Institute of Theoretical Statistics, Copenhagen Business School.
- Bhattacharya, S.: 1979, Imperfect information, dividend policy and the “bird in the hand” fallacy, *The Bell Journal of Economics* **10**, 259–70.
- Bhattacharya, S. and Constantinides, G. M. (eds): 1989, *Theory of valuation: frontiers of modern financial theory*, Vol. 1, Rowman & Littlefield.
- Bierens, H. J. and Broersma, L.: 1993, The relation between unemployment and interest rate: some international evidence, *Econometric Reviews* **12**, 217 – 56.
- Binmore, K. G.: 1977, *Mathematical analysis: a straightforward approach*, Cambridge University Press.
- Black, F.: 1976, The dividend puzzle, *Journal of Portfolio Management* **2**, 72 – 7.
- Black, F. and Scholes, M.: 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* **81**, 637–54.
- Blanchard, O. J. and Fischer, S.: 1989, *Lectures on macroeconomics*, The MIT Press.

- Bollobás, B.: 1990, *Linear analysis: an introductory course*, Cambridge University Press.
- Borowski, E. J. and Borwein, J. M.: 1989, *Collins dictionary of mathematics*.
- Breeden, D. T.: 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7, 265–96.
- Breeden, D. T. and Litzenberger, R. H.: 1978, Prices of state-contingent claims implicit in option prices, *Journal of Business* 51, 621–51.
- Breeden, D. T., Gibbons, M. R. and Litzenberger, R. H.: 1989, Empirical tests of the consumption-oriented CAPM, *Journal of Finance* 44, 231 – 62.
- Brown, D. P.: 1988, The implications of nonmarketable income for consumption-based models of asset pricing, *Journal of Finance* 43, 867–80.
- Brown, D. P. and Gibbons, M. R.: 1985, A simple econometric approach for utility-based asset pricing models, *Journal of Finance* 40, 359–81.
- Brown, S. J., Goetzmann, W. N. and Ross, S. A.: 1995, Survival, *Journal of Finance* 50, 853–73.
- Burnside, C.: 1994, Hansen-Jagannathan bounds as classical tests of asset-pricing models, *Journal of Business & Economic Statistics* 12, 57–79.

- Caballero, R. J.: 1990, Consumption puzzles and precautionary savings, *Journal of Monetary Economics* 25, 113 – 36.
- Caballero, R. J.: 1991, Earnings uncertainty and aggregate wealth accumulation, *American Economic Review* 81, 859 – 71.
- Campbell, J. Y. and Shiller, R. J.: 1988, Dividend-price ratios and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.
- Campbell, J. Y., Lo, A. W. and MacKinlay, C.: Forthcoming, *The econometrics of financial markets*, Princeton University Press.
- Carroll, C. D.: 1992, The buffer-stock theory of saving: some macroeconomic evidence, *Brookings Papers on Economic Activity* 2, 61–135.
- Carroll, C. D.: 1996, Buffer stock saving: some theory, Mimeo: The Johns Hopkins University.
- Carroll, C. D.: 1997, Buffer-stock saving and the life cycle/permanent income hypothesis, *Quarterly Journal of Economics* 112, 1–55.
- Carroll, C. D. and Samwick, A. A.: 1996, The nature of precautionary wealth, Mimeo: The Johns Hopkins University.
- Cecchetti, S. G., Lam, P. and Mark, N. C.: 1990, Mean reversion in equilibrium asset pricing, *American Economic Review* 80, 398–418.

- Cecchetti, S. G., Lam, P. and Mark, N. C.: 1993, The equity premium and the risk-free rate: matching the moments, *Journal of Monetary Economics* **31**, 21–45.
- Cecchetti, S. G., Lam, P. and Mark, N. C.: 1994, Testing volatility restrictions on intertemporal marginal rates of substitution implied by Euler equations and asset returns, *Journal of Finance* **49**, 123–52.
- Chamberlain, G.: 1983, Funds, factors and diversification in arbitrage pricing models, *Econometrica* **51**, 1305–23.
- Chamberlain, G. and Rothschild, M.: 1983, Arbitrage, factor structure and mean-variance analysis on large asset markets, *Econometrica* **51**, 1281–1304.
- Chan, K., Foresi, S. and Lang, L. H. P.: 1996, Does money explain asset returns? Theory and empirical analysis, *Journal of Finance* **51**, 345–61.
- Chen, N., Roll, R. and Ross, S.: 1986, Economic forces and the stock market, *Journal of Business* **59**, 383 – 403.
- Choe, H., Masulis, R. W. and Nanda, V.: 1993, Common stock offerings across the business cycle, *Journal of Empirical Finance* **1**, 3 – 31.
- Cochrane, J. H.: 1991a, Production-based asset pricing and the link between stock returns and economic fluctuations, *Journal of Finance* **46**, 209–37.
- Cochrane, J. H.: 1991b, A simple test of consumption insurance, *Journal of Political Economy* **99**, 957–76.

- Constantinides, G. M.: 1982, Intertemporal asset pricing with heterogeneous consumers and without demand aggregation, *Journal of Business* 55, 253–67.
- Constantinides, G. M.: 1989, Theory of valuation: overview and recent developments, in Bhattacharya and Constantinides (1989).
- Constantinides, G. M.: 1990, Habit formation: a resolution of the equity premium puzzle, *Journal of Political Economy* 98, 519–43.
- Constantinides, G. M. and Duffie, D.: 1996, Asset pricing with heterogeneous consumers, *Journal of Political Economy* 104, 219–40.
- Cox, J. C. and Huang, C.: 1989, Optimal consumption and portfolio policies when asset prices follow a diffusion process, *Journal of Economic Theory* 49, 33–83.
- Cox, J. C., Ingersoll, J. E. and Ross, S. A.: 1985, An intertemporal general equilibrium model of asset prices, *Econometrica* 53, 363–84.
- Cox, J. C., Ross, S. A. and Rubinstein, M.: 1979, Option pricing: a simplified approach, *Journal of Financial Economics* 7, 229–63.
- Dardanoni, V.: 1991, Precautionary savings under income uncertainty: a cross-sectional analysis, *Applied Economics* 23, 153–60.
- Davidson, R. and Mackinnon, J. G.: 1993, *Estimation and inference in econometrics*, Oxford University Press.

Deaton, A.: 1991, Saving and liquidity constraints, *Econometrica* **59**, 1221–48.

Deaton, A.: 1992, *Understanding consumption*, Oxford University Press.

Denis, D. J., Denis, D. K. and Sarin, A.: 1994, The information content of dividend changes: cash flow signaling, overinvestment and dividend clienteles, *Journal of Financial & Quantitative Analysis* **29**, 567–87.

Dimson, E. and Marsh, P.: 1994, UK financial market returns 1953–93, Unpublished manuscript: London Business School.

Dixit, A. K. and Pindyck, R. S.: 1994, *Investment under uncertainty*, Princeton University Press.

Drèze, J. H. and Modigliani, F.: 1972, Consumption decisions under uncertainty, *Journal of Economic Theory* **5**, 308–35.

Duffie, D.: 1988, *Security markets: stochastic models*, Academic Press.

Duffie, D.: 1992a, *Dynamic asset pricing theory*, Princeton University Press.

Duffie, D.: 1992b, The nature of incomplete security markets, in Laffont (1992).

Duffie, D. and Zame, W.: 1989, The consumption-based capital asset pricing model, *Econometrica* **57**, 1279–97.

Dybvig, P. H. and Ross, S. A.: 1992, Arbitrage, *New Palgrave dictionary of money & finance*.

- Dynan, K. E.: 1993, How prudent are consumers?, *Journal of Political Economy* 101, 1104–13.
- Eeckhoudt, L. and Kimball, M.: 1992, Background risk, prudence and the demand for insurance, in G. Dionne (ed.), *Contributions to insurance economics*, Kluwer.
- Eeckhoudt, L., Gollier, C. and Schlesinger, H.: 1993, Changes in background risk, temperance and risk taking behaviour, Unpublished manuscript: HEC School of Management.
- Eeckhoudt, L., Gollier, C. and Schneider, T.: 1995, Risk aversion, prudence and temperance: a unified approach, *Economics Letters* 48, 331–6.
- Epstein, L. G.: 1992, Behaviour under risk: recent developments in theory and practice, in Laffont (1992).
- Epstein, L. G. and Zin, S. E.: 1990, ‘First-order ’ risk aversion and the equity premium puzzle, *Journal of Monetary Economics* 26, 387–407.
- Epstein, L. G. and Zin, S. E.: 1991, Substitution, risk aversion and the temporal behavior of consumption and asset returns: an empirical analysis, *Journal of Political Economy* 99, 263 – 86.
- Fama, E. F.: 1991, Efficient capital markets: II, *Journal of Finance* 46, 1575–617.

- Fama, E. F. and French, K. R.: 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23 – 49.
- Fama, E. F. and Schwert, G. W.: 1977, Human capital and capital market equilibrium, *Journal of Financial Economics* 4, 95–125.
- Feller, W.: 1968, *An introduction to probability theory and its applications*, Vol. One, third edn, John Wiley.
- Ferson, W. E.: 1983, Expectations of real interest rates and aggregate consumption: empirical tests, *Journal of Financial & Quantitative Analysis* 18, 477–97.
- Ferson, W. E.: 1995, Theory and empirical tests of asset pricing models, in R. A. Jarrow, W. T. Ziemba and V. Maksimovic (eds), *Finance*, North-Holland. Handbook of Operations Research & Management Science, Volume 9.
- Ferson, W. E. and Merrick, J. J.: 1987, Non-stationarity and stage-of-the-business-cycle effects in consumption-based asset pricing relations, *Journal of Financial Economics* 18, 127–46.
- Franke, G., Stapleton, R. C. and Subrahmanyam, M. G.: 1994, The size of background risk and the theory of risk bearing, Mimeo: University of Konstanz.

- Franks, J. R., Broyles, J. E. and Carleton, W. T.: 1985, *Corporate finance: concepts and applications*, Kent.
- Friend, I. and Blume, M. E.: 1975, The demand for risky assets, *American Economic Review* 65, 900–22.
- Gaver, J. J. and Gaver, K. M.: 1993, Additional evidence on the association between the investment opportunity set and corporate financing, dividend and compensation policies, *Journal of Accounting & Economics* 16, 125–60.
- Gollier, C. and Pratt, J. W.: 1993, Weak proper risk aversion and the tempering effect of background risk, Unpublished manuscript: HEC School of Management.
- Gollier, C. and Pratt, J. W.: 1996, Risk vulnerability and the tempering effect of background risk, *Econometrica* 64, 1109 – 23.
- Gordon, M.: 1989, Corporate finance under the MM theorems, *Financial Management Summer*, 19–28.
- Gordon, M. J.: 1996, *Finance, investment and macroeconomics*, paperback edn, Edward Elgar.
- Green, H. A. J.: 1976, *Consumer Theory*, revised edn, Macmillan Press.
- Grossman, S. J. and Laroque, G.: 1990, Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods, *Econometrica* 58, 25–51.

- Grossman, S. J. and Shiller, R. J.: 1981, The determinants of the variability of stock market prices, *American Economic Review* **71**, 222–7.
- Grossman, S. J. and Shiller, R. J.: 1982, Consumption correlatedness and risk measurement in economies with non-traded assets and heterogeneous information, *Journal of Financial Economics* **10**, 195–210.
- Grossman, S. J., Mellino, A. and Shiller, R. J.: 1987, Estimating the continuous-time consumption-based asset-pricing model, *Journal of Business & Economic Statistics* **5**, 315 – 27.
- Gruber, J.: 1994, The consumption smoothing benefits of unemployment insurance, National Bureau of Economic Research working paper 4750.
- Guiso, L., Jappelli, T. and Terlizzese, D.: 1992, Earnings uncertainty and precautionary saving, *Journal of Monetary Economics* **30**, 307–37.
- Hakansson, N. H.: 1970, Optimal investment and consumption strategies under risk for a class of utility functions, *Econometrica* **38**, 587 – 607.
- Hakansson, N. H.: 1982, To pay or not to pay dividend, *Journal of Finance* **37**, 415 – 28.
- Hall, R. E.: 1988, Intertemporal substitution in consumption, *Journal of Political Economy* **96**, 339–57.
- Hamilton, J. D.: 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* **57**, 357–84.

- Hamilton, J. D.: 1994, *Time series analysis*, Princeton University Press.
- Hansen, L. P.: 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029 – 53.
- Hansen, L. P. and Jagannathan, R.: 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–62.
- Hansen, L. P. and Singleton, K. J.: 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, 1269 – 86.
- Hansen, L. P. and Singleton, K. J.: 1983, Stochastic consumption, risk aversion and the temporal behavior of asset returns, *Journal of Political Economy* 91, 249 – 65.
- Hansen, L. P. and Singleton, K. J.: 1984, Errata, *Econometrica* 52, 267 – 8.
- Hardouvelis, G. K.: 1988, Economic news, exchange rates and interest rates, *Journal of International Money & Finance* 7, 23 – 35.
- Harrison, J. M. and Kreps, D. M.: 1979, Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* 20, 381–408.
- Harrison, J. M. and Pliska, S. R.: 1981, Martingales and stochastic integrals in the theory of continuous trading, *Stochastic Processes and their Applications* 11, 215–60.

- He, H. and Modest, D. M.: 1992, Market frictions and consumption-based asset pricing, Finance working paper 223, University of California at Berkeley.
- Heaton, J.: 1995, An empirical investigation of asset pricing with temporally dependent preference specifications, *Econometrica* **63**, 681–717.
- Heaton, J. and Lucas, D.: 1992, The affects of incomplete insurance markets and trading costs in a consumption-based asset pricing model, *Journal of Economics, Dynamics & Control* **16**, 601–20.
- Heaton, J. and Lucas, D.: 1993, Evaluating the effects of incomplete markets on risk sharing and asset pricing, National Bureau of Economic Reseach working paper 4249.
- Heaton, J. and Lucas, D.: 1995, The importance of investor heterogeneity and financial market imperfections for the behavior of asset prices, *Carnegie-Rochester Conference Series on Public Policy* **42**, 1–32.
- Hlawitschka, W. and Tucker, M.: 1995, Asset allocation and the equity premium puzzle, *Journal of Business, Finance & Accounting* **22**, 397 – 413.
- Huang, C. and Litzenberger, R. H.: 1988, *Foundations for financial economics*, North-Holland.
- Hubbard, R. G., Skinner, J. and Zeldes, S. P.: 1994, The importance of precautionary motives in explaining individual and aggregate saving, *Carnegie-Rochester Conference Series on Public Policy* **40**, 59–125.

- Huggett, M.: 1993, The risk-free rate in heterogeneous-agent incomplete-insurance economies, *Journal of Economics, Dynamics & Control* **17**, 953–69.
- Hull, J.: 1989, *Options, futures and other derivative securities*, first edn, Prentice-Hall.
- Ingersoll, J. E.: 1987, *Theory of financial decision making*, Rowman and Littlefield.
- Jain, P. C.: 1988, Response of hourly stock prices and trading volumes to economic news, *Journal of Business* **61**, 219–31.
- Jarrow, R. A. and Rosenfeld, E. R.: 1984, Jump risk and the Intertemporal Capital Asset Pricing Model, *Journal of Business* **57**, 337–51.
- Jenkinson, T.: undated, The equity risk premium and the cost of capital debate in the UK regulated utilities, Unpublished manuscript: Oxford University.
- Jensen, M. C.: 1986, Agency costs of free cash flow, corporate finance and takeovers, *American Economic Review* **76**, 323–9.
- John, K. and Williams, J.: 1985, Dividends, dilution and taxes: a signalling equilibrium, *Journal of Finance* **40**, 1053–70.
- Kahn, J. A.: 1990, Moral hazard, imperfect risk-sharing and the behavior of asset returns, *Journal of Monetary Economics* **26**, 27–44.

- Kandel, S. and Stambaugh, R. F.: 1990, Expectations and volatility of consumption and asset returns, *Review of Financial Studies* 3, 207–32.
- Kandel, S. and Stambaugh, R. F.: 1991, Asset returns and intertemporal preferences, *Journal of Monetary Economics* 27, 39–71.
- Kantor, S. E. and Fishback, P. V.: 1996, Precautionary saving, insurance, and the origins of workers' compensation, *Journal of Political Economy* 104, 419–42.
- Karatzas, I., Lehoczky, J. P., Sethi, S. P. and Shreve, S. E.: 1986, Explicit solutions of a general consumption/investment problem, *Mathematics of Operations Research* 11, 261–94.
- Kihlstrom, R. E., Romer, D. and Williams, S.: 1981, Risk aversion with random initial wealth, *Econometrica* 49, 911–20.
- Kimball, M. S.: 1990, Precautionary saving in the small and in the large, *Econometrica* 58, 53–73.
- Kimball, M. S.: 1992, Precautionary motives for holding assets, *New Palgrave dictionary of money & finance*.
- Kimball, M. S.: 1993, Standard risk aversion, *Econometrica* 61, 589–611.
- Kimball, M. S.: 1994, The importance of precautionary motives in explaining individual and aggregate saving: a comment, *Carnegie-Rochester Conference Series on Public Policy* 40, 127–132.

- Kocherlakota, N. R.: 1990a, Disentangling the coefficient of relative risk aversion from the elasticity of intertemporal substitution: an irrelevance result, *Journal of Finance* 45, 175 – 90.
- Kocherlakota, N. R.: 1990b, On tests of representative consumer asset pricing models, *Journal of Monetary Economics* 26, 285 – 304.
- Kocherlakota, N. R.: 1990c, On the discount factor in growth economies, *Journal of Monetary Economics* 25, 43–7.
- Kocherlakota, N. R.: 1996, The equity premium: it's still a puzzle, *Journal of Economic Literature* 34, 42–71.
- Kreps, D. and Porteus, E.: 1978, Temporal resolution of uncertainty and dynamic choice theory, *Econometrica* 46, 185–200.
- Labadie, P.: 1989, Stochastic inflation and the equity premium, *Journal of Monetary Economics* 24, 277–98.
- Laffont, J.-J. (ed.): 1992, *Advances in economic theory*, Vol. 2, Cambridge University Press.
- Lang, L. H. and Litzenberger, R. H.: 1989, Dividend announcements: cash flow signalling vs free cash flow hypothesis?, *Journal of Financial Economics* 24, 181 – 91.
- Layard, R., Nickell, S. and Jackman, R.: 1991, *Unemployment: Macroeconomic performance and the labour market*, Oxford University Press.

- Leland, H. E.: 1968, Saving and uncertainty: the precautionary demand for saving, *Quarterly Journal of Economics* 82, 465–73.
- LeRoy, S. F.: 1973, Risk aversion and the martingale property of stock prices, *International Economic Review* 14, 436–46.
- LeRoy, S. F. and Porter, R. D.: 1981, The present-value relation: tests based on implied variance bounds, *Econometrica* 49, 555–74.
- Lintner, J.: 1956, Distribution of incomes of corporations among dividends, retained earnings and taxes, *American Economic Review* 46, 97–113.
- Lintner, J.: 1965, The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13–37.
- Litzenberger, R. H. and Ronn, E. I.: 1986, A utility-based model of common stock price movements, *Journal of Finance* 41, 67–91.
- Lucas, D. J.: 1994, Asset pricing with undiversifiable income risk and short sales constraints: deepening the equity premium puzzle, *Journal of Monetary Economics* 34, 325–41.
- Lucas, R. E.: 1978, Asset prices in an exchange economy, *Econometrica* 46, 1429–45.
- Luenberger, D. G.: 1969, *Optimization by vector space methods*, John Wiley.

- Mace, B. J.: 1991, Full insurance in the presence of aggregate uncertainty, *Journal of Political Economy* 99, 928–56.
- Mankiw, N. G.: 1986, The equity premium and the concentration of aggregate shocks, *Journal of Financial Economics* 17, 211–9.
- Mankiw, N. G. and Shapiro, M. D.: 1986, Risk and return: consumption beta versus market beta, *Review of Economics and Statistics* 68, 452 – 9.
- Mankiw, N. G. and Zeldes, S. P.: 1991, The consumption of stockholders and nonstockholders, *Journal of Financial Economics* 29, 97–112.
- Marsh, T. A. and Merton, R. C.: 1987, Dividend behavior for the aggregate stock market, *Journal of Business* 60, 1–40.
- Mas-Colell, A., Whinston, M. D. and Green, J. R.: 1995, *Microeconomic Theory*, Oxford University Press.
- Mayers, D.: 1972, Non-marketable assets and capital market equilibrium under uncertainty, in M. C. Jensen (ed.), *Studies in the theory of capital markets*, Praeger.
- Mayers, D.: 1973, Nonmarketable assets and the determination of capital asset prices in the absence of a riskless asset, *Journal of Business* 46, 258–67.
- Mayers, D.: 1976, Nonmarketable assets, market segmentation, and the level of asset prices, *Journal of Financial & Quantitative Analysis* 11, 1–12.

- Mehra, R. and Prescott, E. C.: 1985, The equity premium: a puzzle, *Journal of Monetary Economics* 15, 145–61.
- Mehra, R. and Prescott, E. C.: 1988, The equity premium: a solution?, *Journal of Monetary Economics* 22, 133–36.
- Merton, R. C.: 1969, Lifetime portfolio selection under uncertainty: the continuous-time case, *Review of Economics and Statistics* 51, 247–57.
- Merton, R. C.: 1971, Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory* 3, 373–413.
- Merton, R. C.: 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–87.
- Merton, R. C.: 1992, *Continuous-time finance*, revised edn, Blackwell.
- Miller, M. H.: 1977, Debt and taxes, *Journal of Finance* 32, 261–75.
- Miller, M. H.: 1986, Behavioral rationality in finance: the case of dividends, *Journal of Business* 59, S451–68.
- Miller, M. H. and Modigliani, F.: 1961, Dividend policy, growth and the valuation of shares, *Journal of Business* 34, 411–33.
- Miller, M. H. and Rock, K.: 1985, Dividend policy under asymmetric information, *Journal of Finance* 40, 1031–51.

- Myers, S. C. and Majluf, N.: 1984, Stock issues and investment policy when firms have information that investors do not have, *Journal of Financial Economics* **13**, 187–221.
- Nachman, D. C.: 1982, Preservation of “more risk averse” under expectations, *Journal of Economic Theory* **28**, 361–8.
- Neave, H. R. and Worthington, P.: 1988, *Distribution-free tests*, Unwin Hyman.
- Nelson, J. A.: 1994, On testing for full insurance using consumer expenditure survey data: comment, *Journal of Political Economy* **102**, 384–94.
- Newey, W. K. and West, K. D.: 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**, 703–8.
- Ni, S. X.: 1993, Habit formation as a resolution to the equity premium puzzle: what is in the data, what is not, *Southern Economic Journal* **59**, 749–59.
- Pindyck, R. S.: 1984, Risk, inflation and the stock market, *American Economic Review* **74**, 335–51.
- Poon, S. and Taylor, S.: 1991, Macroeconomic factors and the UK stock market, *Journal of Business, Finance & Accounting* **18**, 619–36.

- Prag, J.: 1994, The response of interest rates to unemployment rate announcements: is there a natural rate of unemployment?, *Journal of Macroeconomics* 16, 171–84.
- Pratt, J. W.: 1964, Risk aversion in the small and in the large, *Econometrica* 32, 122–36.
- Pratt, J. W. and Zeckhauser, R. J.: 1987, Proper risk aversion, *Econometrica* 55, 143–54.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T.: 1989, *Numerical recipes in Pascal: the art of scientific computing*, Cambridge University Press.
- Pruitt, S. W. and Gitman, L. J.: 1991, The interactions between the investment, financing and dividend decisions of major US firms, *The Financial Review* 26, 409–30.
- Rietz, T. A.: 1988, The equity premium: a solution, *Journal of Monetary Economics* 22, 117–31.
- Ross, S. A.: 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–60.
- Ross, S. A.: 1977, Return, risk, and arbitrage, in I. Friend and J. L. Bicksler (eds), *Risk and return in finance*, Vol. 1, Ballinger.
- Ross, S. A.: 1978, A simple approach to the valuation of risky streams, *Journal of Business* 51, 453–75.

- Ross, S. A.: 1981, Some stronger measures of risk aversion in the small and the large with applications, *Econometrica* 49, 621–38.
- Ross, S. A.: 1989, Intertemporal asset pricing, in Bhattacharya and Constantinides (1989).
- Royden, H.: 1968, *Real analysis*, second edn, Macmillan.
- Rubinstein, M.: 1974, An aggregation theorem for securities markets, *Journal of Financial Economics* 1, 225–44.
- Rubinstein, M.: 1976, The valuation of uncertain income streams and the pricing of options, *The Bell Journal of Economics* 7, 407 – 25.
- Sadeghi, M.: 1992, Stock market response to macroeconomic news: the Australian evidence, International Monetary Fund Working Paper WP/92/61.
- Saito, M.: 1993, Incomplete insurance and risk-free rates: the case of permanent idiosyncratic shocks, Working paper: University of British Columbia.
- Sandmo, A.: 1970, The effect of uncertainty on saving decisions, *Review of Economic Studies* 33, 353–60.
- Sant, R. and Cowan, A. R.: 1994, Do dividends signal earnings? the case of omitted dividends, *Journal of Banking & Finance* 18, 1113–33.

- Sauer, A. and Murphy, A.: 1992, An empirical comparison of alternative models of capital asset pricing in Germany, *Journal of Banking & Finance* 16, 183 – 96.
- Scheinkman, J. A.: 1989, Market incompleteness and the equilibrium valuation of assets, in Bhattacharya and Constantinides (1989).
- Scott, M. F.: 1992, The cost of equity capital and the risk premium on equities, *Applied Financial Economics* 2, 21–32.
- Scott, M. F.: 1993, Real interest rates: past and future, *National Institute Economic Review* February, 54–71.
- Sharpe, W. F.: 1964, Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–42.
- Shiller, R. J.: 1981, Do stock prices move too much to be justified by subsequent changes in dividends?, *American Economic Review* 71, 421–36.
- Siegel, J. J.: 1992, The equity premium: stock and bond returns since 1802, *Financial Analysts Journal* January - February, 28 – 38.
- Siegel, J. J. and Thaler, R. H.: 1997, The equity premium puzzle, *Journal of Economic Perspectives* 11, 191–200.
- Singleton, K. J.: 1990, Specification and estimation of intertemporal asset pricing models, in B. Friedman and F. Hahn (eds), *Handbook of monetary economics*, Vol. 1, Elsevier Science.

- Skinner, J.: 1988, Risky income, life cycle consumption and precautionary savings, *Journal of Monetary Economics* **22**, 237 – 55.
- Smith, C. W. and Watts, R. L.: 1992, The investment opportunity set and corporate financing, dividend and compensation policies, *Journal of Financial Economics* **32**, 263–92.
- Snow, K. N.: 1991, Diagnosing asset pricing models using the distribution of asset returns, *Journal of Finance* **46**, 955–83.
- Stapleton, R. C. and Subrahmanyam, M. G.: 1978, A multiperiod equilibrium asset pricing model, *Econometrica* **46**, 1077–96.
- Stapleton, R. C. and Subrahmanyam, M. G.: 1979, Marketability of assets and the price of risk, *Journal of Financial & Quantitative Analysis* **14**, 1–10.
- Svensson, L. E. and Werner, I. M.: 1993, Nontraded assets in incomplete markets: pricing and portfolio choice, *European Economic Review* **37**, 1149–68.
- Telmer, C. I.: 1993, Asset-pricing puzzles and incomplete markets, *Journal of Finance* **48**, 1803–32.
- Weil, P.: 1989, The equity premium puzzle and the risk-free rate puzzle, *Journal of Monetary Economics* **24**, 401 – 21.
- Weil, P.: 1992a, Equilibrium asset prices with undiversifiable labor income risk, *Journal of Economics, Dynamics & Control* **16**, 769–90.

- Weil, P.: 1992b, Hand-to-mouth consumers and asset prices, *European Economic Review* **36**, 575–83.
- Weil, P.: 1994, Nontraded assets and the CAPM, *European Economic Review* **38**, 913–22.
- Weston, J. F.: 1989, What MM have wrought, *Financial Management Summer*, 29–38.
- Wheatley, S.: 1988, Some tests of the consumption-based asset pricing model, *Journal of Monetary Economics* **22**, 193 – 215.
- Williams, D.: 1991, *Probability with martingales*, Cambridge University Press.
- Yoon, P. S. and Starks, L. T.: 1995, Signaling, investment opportunities and dividend announcements, *Review of Financial Studies* **8**, 995–1018.
- Zeldes, S. P.: 1989a, Consumption and liquidity constraints: an empirical investigation, *Journal of Political Economy* **97**, 305–46.
- Zeldes, S. P.: 1989b, Optimal consumption with stochastic income: deviations from certainty equivalence, *Quarterly Journal of Economics* **104**, 275–98.
- Zin, S. E.: 1995, The importance of investor heterogeneity and financial market imperfections for the behavior of asset prices. a comment., *Carnegie-Rochester Conference Series on Public Policy* **42**, 33–8.